36 No. 1 (New Series No. 169) October 1975

ANALYSIS

Edited by

C. J. F. WILLIAMS

CONTENTS

BASIL BLACKWELL . ALFRED STREET . OXFORD

Price £1.20

ANALYSIS 36.1 OCTOBER 1975

EDITORIAL

TO save costs the ANALYSIS Committee has agreed to reduce the number of issues per volume to four while retaining the overall number of pages in a volume. Three issues of 48 pages will appear in October, March and June, respectively, and one of 64 pages in January. Printing difficulties have made prompt appearance at the old times practically unattainable, so it is hoped that readers will not feel that much of value has been lost.

DUMMETT ON ASSERTION

By MICHAEL COHEN

IN his article 'Truth' Michael Dummett makes the following criticism of Frege's account of truth:

> . . . it is part of the concept of truth that we aim at making true statements; and Frege's theory of truth and falsity as the references of sentences leaves this feature of the concept of truth quite out of account. Frege indeed tried to bring it in afterwards, in his theory of assertion—but too late; for the sense of a sentence is not given in advance of our going in for the activity of assertion, since otherwise there could be people who expressed the same thoughts but went in instead for denying them.[1]

And this supposition—that, as Wittgenstein puts it at *Tractatus* 4.062, we could

> make ourselves understood with false propositions just as we have done up till now with true ones . . . so long as it's known that they are meant to be false

—is, as he points out, absurd.

> For a proposition is true if things are as we use it to say they are; and if by '*p*' we mean $\sim p$ and things are as we mean, then on the new interpretation '*p*' is true and not false.

[1] M. Dummett, 'Truth' in *Truth*, ed. G. Pitcher, p. 95.

And so is no longer 'the same proposition' of course; just the same sign with a new sense. The point here is not, as Professor Anscombe would have it, that 'A code by which one always meant the negative of what one said need not break down'[1]; it is rather that there could be no language whose speakers always meant the negative of what they said.

Frege had thought that the notions of sense and reference were inadequate as an account of what it is to speak a language. One might utter a sentence (thereby expressing a thought and referring to a truth-value) without committing oneself to the truth of the sentence—as for example when I utter the sentence 'It is raining' in the context of the complex sentence 'If it is raining, the streets will be wet'. One needs a sign, Frege thought, to indicate those utterances in which one does commit oneself to the truth of a sentence. Now from the fact that just to utter a sentence is not *always* to commit oneself to its truth it does not follow that it *never* is; indeed the latter idea is incoherent, which is why Wittgenstein says at *Tractatus* 4.442 that Frege's assertion sign is 'logically quite meaningless'. But the main point is not that Frege thought that a sign was necessary to indicate assertion, but that even if just to utter a sentence *could* be to commit oneself to its truth (as he seems later to have thought), this could be no more, given his account of truth, than an arbitrary convention.

In his recent book on Frege, Dummett has tried to defend Frege against this form of attack by arguing that what Wittgenstein (and he) had thought to be incoherent is not a possibility on Frege's theory. He agrees though with Frege that

> The notions of sense and reference do not suffice for a complete account of language. If we know of a language only what sense the expressions which occur in it have, and thereby their reference, we know nothing which can tell us the significance of uttering an expression of this language: the *point* of doing so.[2]

Dummett compares Frege's theory of sense and reference—in its application to sentences—to the formal description of a game like chess. The formal description of chess consists of a description of the initial and end positions of the game, together with an account of which moves are permissible in each position in the game. It is clear that one could grasp the formal description of chess without knowing what it is to play *chess*, since there are other games which share that formal description; for example the game in which each player's aim is to force his opponent to checkmate him. To turn the formal description into the description of a particular game—chess, say—we need to add a specification of certain positions as winning ones. If we know that the point of playing is to win we should, given this specification, know what it is to play chess.

[1] G. E. M. Anscombe, *An Introduction to Wittgenstein's Tractatus*, p. 77.
[2] M. Dummett, *Frege: Philosophy of Language*, p. 295.

That the point of playing is to win is, Dummett supposes, a remark parallel to the remark that in speaking we aim at saying what is true. But this makes nonsense of the analogy; for it suggests that what one is ignorant of if one knows only the sense and reference of sentences is: which of the sentences are true. But for Frege the reference of a sentence *is* its truth-value. And it won't do to say that knowing the formal description of a game is the analogue of knowing just the sense (and not the reference) of a sentence; for the implication here is that the same sense might be true in one language and false in another (as the same position might be a winning one in one game and a losing one in another).

Elsewhere Dummett speaks as though what is missing from an account of language in terms of the sense and reference of its sentences is the *significance* of the distinction between truth and falsity.

> . . . just as it is possible to describe to someone what it is to play chess without presupposing that he understands what winning is, and therefore that he understands some similar activity, so it should be possible to describe the activity of using language without presupposing that it is already known what significance it has to call one class of sentences the class of 'true' sentences and the other the class of 'false' sentences; this feature, to be expressly described, is not contained in the characterization of our language in terms of sense and reference (ibid., p. 297).

It is clear that on this way of taking the analogy knowing the sense and reference of sentences is no longer thought of as the analogue of knowing the formal description of a game, in the sense in which this is a common feature of various games. What one might fail to know about chess is not which positions are called 'winning' but what it means to call them this; but chess differs from its variants precisely in that it involves different winning positions from them.

One difficulty with the analogy between speaking and playing a game is in knowing what Dummett means by an 'account' of language. We might say that someone fails to understand chess (or equally that he fails to understand an account of it) if, while knowing which positions are called 'winning', he does not know that this expression marks out those positions to be aimed at in playing. But it could not be like this with language. In the first place, we learn language without learning which sentences are true and false; but the game analogy has no place for the distinction between the sense and reference of sentences—which distinction is of course Frege's way of explaining how we can understand a sentence without knowing its truth-value. To know a game is to know what it is to win that game. And secondly it is absurd to suppose that there might be an 'account' of language for beings who had no 'prior knowledge of some activity akin to human language'; as if such beings might learn to 'describe' our activities without themselves having a language and so the distinction between truth and falsity.

Elsewhere though, Dummett speaks as though an 'account' of a language in terms of the sense and reference of its sentences is something like the description we might give of a foreign language. The inadequacy of this account would lie in the fact that we might be said to know the sense (and reference) of the sentences of this language without knowing what purpose its speakers have in uttering them; without knowing, that is, whether they mean themselves to be taken as saying what is true or what is false. (Dummett slightly obscures this issue by supposing that the alternative hypotheses about their purpose are hypotheses about the meaning of a particular sign in their language, a sign which, prefixed to a sentence, might indicate that that sentence is to be taken as expressing a truth, or a falsehood. This is rather like supposing that the expressions 'win' and 'lose' are counters in a game, rather than parts of the game's description.)

It is quite unclear from what he says whether Dummett imagines the possibility that there is a language whose speakers intend themselves to be taken as saying what is false is a real one or not. At times he explicitly denies that it is. Elsewhere though he says that this hypothesis would be only 'spuriously distinct' from the hypothesis that the speakers of such a language mean to be taken as saying what is true, since we can always adjust the account we give of the senses of their sentences to agree with either hypothesis. If we discover in this language a sentence which is used in the sort of situation in which we should say 'It is raining' we might suppose either that this sentence means that it's raining and that they mean to be taken as saying what is true; or that the sentence means that it's dry and that they mean to be taken as saying what is false. But no observation of how they use this sentence could decide that one hypothesis rather than the other is right; so the hypotheses are only 'spuriously distinct'.

In view of this claim it is quite amazing that Dummett goes on to say that since

> it is indifferent which we choose, it is clear that we shall always choose the one which resembles the account we give of our own language most closely, namely the interpretation under which . . . their intention in speaking is to say what is true (ibid., p. 319).

If the two hypotheses really are indistinguishable, we might as well describe our own practice as 'meaning what we say to be taken as false' . . . provided we suitably adjust the 'account' we give of the senses of our own sentences! Dummett quite misses the point of Wittgenstein's remarks here, for it is just the 'indistinguishability' of the two hypotheses that leads to the question 'Could we not make ourselves understood with false propositions . . . ?' that Wittgenstein rejects as absurd.

Dummett cannot have it both ways. If it is no mere convention that the speakers of a language mean to be taken as saying what is true, then

the account of a language in terms of the senses of its sentences is adequate, for we add nothing to that account with the claim that this *is* their intention. To say on the other hand that it makes no difference whether it be said that their intention is to be taken as saying what is true or as saying what is false is to make the notion of sense not inadequate but futile: we never could know the sense (nor, therefore, the truth-value) of a sentence of their language.

University College of North Wales, Bangor

A NOTE ON THE CONJUNCTIVITY OF KNOWLEDGE

By ALAN H. GOLDMAN

THE lottery paradox is by now a familiar topic in discussions of rational belief. (It was introduced by Kyburg in [5], p. 197.) It results from acceptance of three principles: (1) a probabilistic rule of acceptance—a rational agent will believe any hypothesis with a high enough probability of being true (relative to his evidence), even if the probability is less than 1; (2) a consistency principle—the set S of rational beliefs will not contain contradictory members; (3) a deductive closure principle—the set S of rational beliefs contains the logical consequences of any of its subsets; and following from the last, a principle of conjunctivity—the set S of rational beliefs contains the conjunction of any finite number of its members. (1) would have us believe of each ticket in a fair lottery with enough tickets that it will not win; (3) would then have us believe that no ticket will win, which violates (2) in being inconsistent with the belief that the lottery is fair and hence that some ticket will win. In fact we need only the probability calculus to know that if we accept or believe hypotheses which we recognize to have a probability less than 1 (any belief acquired by a fallible method might be so viewed), a conjunction of such beliefs will, given enough members, have a probability less than 0·5.

The strategies for avoiding this unhappy result involve either a denial of deductive closure or, more specifically, the conjunctivity principle (consequences of individual beliefs may still be allowed in the set S) (Kyburg); or in this case a denial that one would rationally accept any of the hypotheses as to the individual tickets' not winning. Various options will imply the latter: (a) having no acceptance or detachment rule at all, but only partial beliefs, beliefs in probabilities—we believe

here not that each ticket will not win, but only that its probability of winning is $1/n$ (Jeffrey); (b) having no probabilistic rule of acceptance—no matter how high we take the probability of a given hypothesis to be, this in itself does not give us reason to accept it (Harman); (c) having no purely probabilistic rule of acceptance, but only local acceptance of hypotheses more probable than other competing hypotheses relative to some ultimate partition relative to a given case or problem—here the competing hypotheses would refer to the chances of each ticket; all being equiprobable, belief would be suspended (Levi, Lehrer).

My own view is that Kyburg's initial suggestion to give up conjunctivity for rational belief is best. The problem with alternatives (a) and (b) is that they are false to ordinary usage and psychology. We do believe certain things and disbelieve others, although most of these beliefs are rightly taken to be fallible, and we are likely to believe what we take to have an overwhelming likelihood of being true. Furthermore, we know only what we believe or accept: tentative or partial belief is not sufficient for knowledge. Finally, we act on our set of beliefs as a current model of the world, although we do not believe this model infallible or even entirely true. Clinging to the logical principle of conjunctivity at the price of condemning such common practices and usages is relegating one's model of rational belief unnecessarily to the realm of philosophical idealization.

The problem with alternative (c), as Kyburg has argued (in [4], pp. 69–70), is that it collapses into a denial of conjunctivity, since acceptance is relative to partition (relevant hypotheses) for a given problem, and we may not conjoin beliefs based upon different partitions. If $h_1 \ldots h_n$ represent hypotheses that ticket$_1 \ldots _n$ will win, we do not accept $\sim h_1$ vs $\sim h_2$, etc., but we do accept $\sim h_1$ vs $\sim h_2$ & $\sim h_3$ & $\sim h_4$. Given n partitions representing the chances of each ticket against all the others, we get beliefs that each ticket will not win, and the prohibition from conjunction becomes indistinguishable from Kyburg's own solution.

I will not argue further here directly against the conjunctivity of rational belief, since I will attack a principle apparently more obvious, and if I am right that this latter principle is wrong, we should be far less reluctant to abandon the former. In any case, the fact that the logic of belief, even rational belief, does not meet principles of truth-functional deductive logic should no longer surprise us.

I turn to the principle dubbed the Conjunctivity of Knowledge by D. M. Armstrong (in [1], pp. 185–7). He accepts the rejection of the conjunctivity of rational belief, but holds that if we replace belief with knowledge in the antecedent of the principle, it becomes valid. Thus while it is not the case that A's rational belief in each of a set of propositions makes it rational for him to believe their conjunction, it is the case,

Armstrong holds, that if A knows each of a set of propositions, then it is rational for him to believe their conjunction. $\sim((RBAp \,\&\, RBAq \ldots) \supset RBA(p \,\&\, q \ldots))$, but $(KAp \,\&\, KAq \ldots) \supset RBA(p \,\&\, q \ldots)$. Applied to the lottery example, this new principle seems plausible: while rational belief in each ticket losing does not make it rational to believe no ticket will win, knowledge that each ticket will lose certainly seems to imply that no ticket will win, that the lottery is fixed. The latter implication seems to hold simply because the truth of h is one of the criteria for knowledge of h, and if it is true of each ticket that it will lose, then it is true that all will lose.

Despite its plausibility when applied to the lottery case, I believe we must reject the Conjunctivity of Knowledge, and not merely the principle as applied to rational belief. To see why the former fails, we must turn briefly to familiar Gettier cases regarding knowledge as justified true belief. These result from the joint principles (A) that one can be justified in believing a false proposition, and (B) that if one is justified in believing p and infers q from p, then he is justified in believing q. If one is justified in believing a false p, from which he infers a true q, he will have justified true belief on these principles which is not knowledge. While Gettier took these cases to defeat the traditional analysis of knowledge, several philosophers have recently argued (e.g., Meyers and Stern, [9]), rightly I believe, that principle (B) above should rather be rejected, allowing us to retain the traditional analysis by interpreting the justification criterion so as to avoid Gettier counterexamples.

Whatever justification there may be for the rejection of (B), however, no philosopher I know of has rejected (A), and my argument against the Conjunctivity of Knowledge (to return to the main topic) rests upon its incompatibility with that principle, with the acceptance of fallible justificatory sources or with justified false beliefs, together with the retention of the traditional analysis of knowledge as outlined above. Rejection of (A) seems all but impossible. Examples of justified false beliefs abound in the literature. To cite just one, if Jones, a usually trustworthy friend, drives up in a Ford of the same type I know him to have owned yesterday and refers to it as his car, I am justified in believing it to be his car, whether in fact he is deceiving me or not. To claim that I am not justified in believing it to be his car if that turns out false is to imply that I do not know it to be his if in fact it is, since my evidence is the same in the two cases, and I can be justified only in relation to the actual evidence (given that the evidence, including intermediate inferences from it, is true). But it is clear that we would not and should not hesitate to say of the truthful Jones that I know him to own that Ford. What better evidence could I demand for the justification of my belief (the production of a registration could also be phony)?

But retention of (A) is incompatible with the Conjunctivity of Know-

ledge. The possibility of justified false belief shows that it is possible for the only difference between a proposition known and one merely believed to be that the former is true and the latter false (cf. Malcolm, [8], pp. 58–60). If this is so, then it is possible to have two sets of propositions, p, q, . . . and p_1, q_1, . . . such that each proposition in both sets has equal justification, but in the first set several or all are false, and in the second all turn out true. It will then be the case if A believes all, that several or all of p, q, . . . are merely believed, while all of p_1, q_1 . . . are known. If, regarding the former, $JBAp$ & $JBAq$ (A is justified in believing p, etc.) . . . does not entail RBA (p & q . . .), and $RBAp$ & $RBAq$ . . . does not entail RBA (p & q . . .),[1] then I believe we must hold regarding the second set that KAp_1 & KAq_1 . . . does not entail RBA (p_1 & q_1 . . .), since the justification and rationality of belief in the propositions of both sets is by hypothesis the same.

The incompatibility of (A) and the Conjunctivity of Knowledge can be seen less abstractly by example. Suppose A has to total 500 sets of figures and does so by the usual method of non-machine-aided addition. If all his totals turn out correct, we have no hesitation in saying that he knows each, since his method of arriving at belief was reliable and hence justificatory. But given that the average totaler, which A is, makes 1 error in every 200 additions, it would not be rational for him to believe that he has made no error, that all his totals are the real sums. He knows each but should not believe their conjunction, contrary to the principle in question.

Generalizing again from this example, suppose that a certain source or method is reliable in generating true beliefs $x\%$ ($95 < x < 100$) of the time and accordingly can be appealed to in order to justify beliefs which it generates, i.e. confers the status of knowledge upon beliefs when they are true. Suppose further that $x+n$ of A's beliefs generated by this method are true, that he therefore not only believes but knows in each case. If he also knows the degree of reliability of his fallible source ($x\%$), e.g. his own perception or memory, it nevertheless will not be rational for him to believe their conjunction, i.e. that all $x+n$ are in fact true. This defeats the conjunctivity principle.

The above example shows Armstrong's principle to lack even the intuitive support he accords it. Although in certain cases, depending on the means of justification and specifically on whether several independent means are available, knowledge of each of a set of propositions indicates the rationality of belief in their conjunction, just as for simple propositions on a given subject matter knowledge of each points to knowledge of their conjunction, there is no entailment in the former case

[1] Justification implies rationality but is not implied by it, given rejection of (B) but acceptance of a probabilistic rule of rational acceptance. The lottery paradox, however, works for justification as well as rationality, (B) being irrelevant to it; hence the rejection of conjunctivity for justification.

just as there is not in the latter. It turns out that it is only when A knows that he knows each of a set of propositions that it is entailed that it is rational for him to believe their conjunction. $KA(KAp \,\&\, KAq \ldots) \supset RBA(p \,\&\, q \ldots)$, but $\sim((KAp \,\&\, KAq \ldots) \supset RBA(p \,\&\, q \ldots))$. (The antecedent of the former is also strong evidence that A knows the conjunction $(p \,\&\, q \ldots)$.) My argument therefore presupposes further that A can know each of a set of propositions without knowing that he knows each, which I accept. I take it that this means only that he has no independent method, no method other than the fallible one in question, for verifying the truth of his beliefs. What other import can knowing that one knows have than justifying one's belief that one knows by an independent method?

While the plausibility of the Conjunctivity of Knowledge then results from the truth criterion of knowledge considered alone, its rejection rests upon the fact that knowledge can involve fallible (even if always true) belief, and upon the fact that A may know that p without knowing that he knows it. One upshot of the rejection of this principle is, as I indicated near the beginning, that Kyburg's own solution to his lottery paradox appears superior to those of his critics. The conjunctivity of rational belief appears far less plausible to the degree that its stronger counterpart fails. Certainly no one could accept the principle as applied to rational belief alone and reject it as applied to knowledge (since knowledge entails rational belief). Another result is that we can continue to accept fallible sources of beliefs as justificatory, as sources of knowledge, as we could not if the Conjunctivity of Knowledge were valid. My argument has been based upon the relative centrality of the former in our conceptual scheme of knowledge and evidence. It seems perfectly obvious to me that I know there is a desk in the next room even if statistically such judgments sometimes turn out to be wrong and thus have a probability less than 1.

References

[1] Armstrong, D. M., *Belief, Truth and Knowledge* (Cambridge: 1973).
[2] Harman, G. H., *Thought* (Princeton: 1973).
[3] Jeffrey, R. C., *The Logic of Decision* (New York: 1965).
[4] Kyburg, H. E., 'Conjunctivitis', in *Induction, Acceptance, and Rational Belief,* edited by M. Swain (Dordrecht, Holland: 1970).
[5] ——, *Probability and the Logic of Rational Belief* (Middletown: 1961).
[6] Lehrer, K., 'Justification, Explanation, and Induction', in *Induction, Acceptance, and Rational Belief,* edited by M. Swain (Dordrecht, Holland: 1970).
[7] Levi, I., *Gambling with Truth* (New York: 1967).
[8] Malcolm, N., *Knowledge and Certainty* (New Jersey: 1963).
[9] Meyers, R. G., and Stern, K., 'Knowledge without Paradox', *The Journal of Philosophy,* LXX (March 22, 1965).

University of Idaho

SCHEMATIC LETTERS AND VARIABLES

By V. H. DUDMAN

QUINE'S distinction between schematic letters and the bindable variables of quantification has recently been challenged by Dummett.[1] Variables take values while schematic letters are merely assigned *interpretations*; but, Dummett claims, the distinction is ontologically baseless in the long run, because interpreting a schematic letter comes down finally to associating with it a range of appropriate *entities*, whereupon schematic letters are seen to be no less ontologically loaded than Quine takes variables to be. By an accident of context, some of Dummett's remarks are cast in terms of a notion of reference which he draws from Frege. However, since his argument against Quine does not (and cannot) turn upon this notion of reference, I shall steer clear of it in what follows, where I try to show that Dummett fails to establish his case.

Obviously, to say that '*p*', '*q*', etc., dummy for indicative sentences and that '*F*', '*G*', etc., dummy for predicates is insufficient to clear these letters of ontic responsibility, for the question then simply devolves upon that of the semantic roles of sentences and predicates themselves: everything depends on whether sentences and predicates are ontologically innocent. And it is Dummett's contention that they are not. When we come to explain the semantic roles of these expressions, we have to explain that 'a constituent sentence contributes to determining the truth-value of the complex sentence of which it is a part in virtue of its truth-value' and that 'a one-place predicate so contributes in virtue of the set of objects to which it applies; and so on' [522]. Such explanations do not indeed require the construal of sentences as *names* of their truth values or predicates as *names* of their extensions—which is just as well, since sentences and predicates are not names of anything. But naming, Dummett urges, is beside the ontological point. The mere fact that our semantic account of language involves the setting up of *associations* between sentences and truth values, between predicates and sets, is enough to convict sentences and predicates (and hence their schematic letters) of ontological import, for the *entities*, the truth values and sets, 'have still been "countenanced" or admitted to the ontology' [525].

The problem about this argument for me is that I doubt its premisses: I am unable to see that Quine is compelled to set up the associations Dummett speaks of. My problem is exacerbated by Dummett's contention [525] that 'certainly it is in no way apparent that he has any alternative type of semantic account to offer', for Quine's general semantic theory, as I apprehend it, contains very definite alternative doctrines.

[1] Michael Dummett, *Frege: Philosophy of Language*, London, 1973.

First, as regards sentences. There are no such things as truth values. Truthfunctionally complex sentences are true or false depending, in one way or another, on whether their constituent sentences are true or false. But we can give an exhaustive account of such dependence without invoking truth values as entities ('every conjunction of true sentences is true', etc.). Certainly the substantive 'truth value' provides an immensely convenient *façon de parler*. Equally certainly, though, it is a prime target for elimination by paraphrase: talk about truth values can be replaced by talk about true and false sentences at the cost only of style, so that we need not give even the appearance of "countenancing" truth values.

The case of predicates is slightly different, since Quine does not of course deny the existence of sets in general or, in particular, of sets as the extensions of predicates. What he does deny, on my understanding of his position, is this: that a satisfactory account of the semantic role of predicates necessarily includes the claim that predicates have sets associated with them as their extensions. It is indeed a convenience to associate with each predicate the set of all and only those objects of which it is true, for then our model theory can enjoy the comforts of set theory. But in taking this step—for the theory of predication a gratuitous step—we are going far beyond the task of describing the semantic role of predicates in our language. It is the role of the (one-place) predicate in predication to be true of each severally of any number of objects [cf. *Word and Object,* §§19, 20].

Are we shrugging off our ontic commitments too easily with these sketchy denials? Let us not forget that, in our model theory, we quantify over the interpretations of schematic letters: when it comes to defining logical consequence, satisfiability and the like, we have to invoke the notion of truth under every interpretation [thus Dummett, 526]. Can we even here persist in disclaiming truth values and sets as associated entities? Yes, Quine has argued: the interpretations of sentence-letters and predicate-letters over which we are quantifying can be taken to be respectively sentences and predicates themselves. We can understand 'comes out true under every interpretation of its schematic letters' as meaning 'comes out true under every replacement of its sentence-letters by actual sentences (of the object-language in hand) and under every replacement of its predicate-letters by actual predicates (of that language)'. We quantify thus over expressions, continuing the while to deny the need for entities associated with them. In the case of sentences, it is no great strain to accept that 'truth under all substitutions' accomplishes everything that 'truth under every assignment of truth values' might achieve, for 'even the poorest language has a true statement and a false one' [*Methods of Logic* 2nd ed., 95]. (I take Excluded Middle as common ground in the present context: cf. 527.) That the parallel result holds in the case of predicates is less evident, and time was when Quine

guarded against a discrepancy by understanding 'true under every interpretation' as 'true for every class as extension' [*loc. cit.*]. But he has since argued that there was no need: the substitutional definition delivers the same goods—at any rate for all object-languages rich enough for elementary number theory [*Philosophy of Logic,* 53–5]. (Even in the case where we choose all our interpretations from some back-woods object-language which does not run to elementary number theory, any resultant divergence between, say, logical consequence, thus defined, for that language and "genuine" logical consequence will lead to no trouble *within* that language.) At the level of metalinguistic discourse, in statements of implication and so on, sets admittedly get in on the act (though still not necessarily as entities "associated" with object-language predicates); for at this level we are talking about, quantifying over, object-language expressions, and these expressions are presumably sets. But that is another point entirely.

Does he who affirms in so many words that Socrates is a man thereby explicitly commit himself to the existence of a particular object, a particular set (or concept) and a particular truth value? I am not sure how Dummett would respond, chiefly because I am not sure why he adopts the position I have been discussing. As far as I can see, he advances no reason in support of his claim that we cannot explain the semantic roles of sentences and predicates in general without associating ranges of entities with them. In the absence of such reasons it is "in no way apparent" that Quine's sharp distinction between schematic letters and bindable variables is, as Dummett puts it [522], 'a misconstrual of the situation'.

Not unnaturally, Dummett contemplates the binding by quantifiers of letters in sentence position and letters in predicate position, as well as of those in singular term position. One cannot but admire the egregious grammatical versatility of Dummett's quantifier, an object-language symbol which contrives to make equally good sense when combined, in exactly the same way, with expressions belonging to three utterly disparate grammatical categories.

Macquarie University

IFS, HOOKS AND ILLOCUTIONARY ACTS

By VERA PEETZ

IN this paper, I propose to look at the more usual *ifs* in the context of various types of illocutionary act and to consider whether any of these *ifs* correspond to material implication. There is one sort of *if* which I will not discuss—the contrary-to-fact conditional.

Before going on to discuss *ifs* in the context of illocutionary utterances, I would like to say something about Austin's *ifs*. Austin[1] distinguished three types of *if*: (1) the normal conditional (or causal) *if*; (2) the stipulative *if*; (3) the *if* of doubt or hesitation. Austin's third *if*, the *if* of doubt or hesitation, appears in an antecedent *if* clause with no accompanying consequent clause, and is attached usually to some assertion, for example, 'There are biscuits on the table if you want them'. However, it has been argued (by Professor Fogelin,[2] for example) that a companion consequent clause can be provided for the solitary *if* clause, that in fact 'There are biscuits on the table if you want them' is really an ellipsis for 'There are biscuits on the table. If you want some biscuits, have some of the biscuits on the table.' (Perhaps the reason why such *ifs* are abbreviated in the way they are is that the fully explicit version is so repetitive and people are very lazy about being fully explicit if they can make their meaning clear by a shorter cut.) We might say that though there are no *ifs* of doubt or hesitation, there are degenerate *ifs*: *ifs* such as 'If you like!', 'If you don't mind!', 'If it isn't Agatha!'. Such *ifs* could be supplied with suitable consequent clauses, but their real function is no longer to express a condition but an attitude.

Attempts have also been made to reduce Austin's stipulative *ifs* to normal conditional (or causal) *ifs* (for example, by Professor Fogelin and Mr. Don Locke[3]), but this reduction cannot be made. Austin pointed out that a normal conditional *if* must satisfy two conditions: (a) that it contraposes and (b) that from it one cannot infer the consequent *simpliciter*. (In the example Austin gives, from 'If I run, I pant' one can infer 'If I do not pant, I do not run', but one cannot infer 'I pant' *simpliciter*.) These conditions, said Austin, were not satisfied by stipulative *ifs* such as 'I promise to marry him if he asks me', since we cannot contrapose this to 'If I do not promise to marry him, he does not ask me' and one can infer *simpliciter* that I promised. However, this will not do. The example of a normal conditional *if* to which Austin applies his criteria ('If I run I pant') is an illocutionary act whose illocutionary force has not been made explicit; that is, it is, in effect, the content of an

[1] *Philosophical Papers*, Oxford, 1961.
[2] See 'Austinian Ifs', *Mind*, 1972.
[3] See 'Ifs and Cans Revisited', *Philosophy*, 1962.

illocutionary act (the fully explicit act being presumably something like 'I assert that if I run I pant'). On the other hand, the example which Austin gives of a stipulative *if* is, in fact, a fully explicit illocutionary act, and to it, illocutionary force bearing part included, Austin applies his two criteria. But Austin cannot apply his criteria differently in each case. The proper procedure would seem to be to apply Austin's two criteria to the content only, since the illocutionary force bearing part stands outside the *if*-containing content. Applying this to Austin's example of a stipulative *if*, we see that 'I promise to marry him if he asks me' does not satisfy the first condition, for we cannot have 'I promise that he does not ask me, if I do not marry him', but that it does satisfy the second condition, since if I say 'I promise to marry him if he asks me', one cannot infer that I promised to marry him *simpliciter*.

Austin's conditions will have to be revised thus: (1) In an illocutionary act, an *if* in the content is a normal conditional (or causal) *if*, if it contraposes within the context of the act. If it is a stipulative *if*, it does not contrapose. (2) With both a normal conditional (or causal) *if* and a stipulative *if*, if the utterance is explicitly 'I *R* that if *a* then *b*' (where '*R*' is the illocutionary force bearing verb, bearing its illocutionary force), it cannot be inferred that I *r*-ed *b* (where '*r*' is the illocutionary force bearing verb, not bearing its illocutionary force and being used descriptively).

Let us now consider *ifs* in the context of speech acts. In a previous article ('Propositional Content', ANALYSIS 32.6), I distinguished between two types of illocutionary act: those whose contents are propositions and those whose contents are not propositions. These two types are illustrated respectively in groups A and B.

A.
(1) I state that if I run I pant
(2) I predict that if it rains the match will be cancelled
(3) I guess that if you are holding a card in your right hand it is the ace of spades

B.
(4) I promise to reward you if you find my book
(5) I order you to return it if you find it
(6) I warn you to be careful if you see the bull

In group A the contents of the acts are propositions. In group B the contents of the acts are not propositions (even though they might be put in apparently propositional form, for example, 'I promise that I will reward you if you find my book'—the 'will' here is the 'will' of determination and willing rather than the future-tensed 'will'.) Now it might be said that if the contents of the speech acts in group B are not propositions there is no point in discussing whether the *ifs* in them could be equivalent to material implication, since it is propositions that are con-

nected by the hook of material implication. There are two ways in which we might meet this difficulty. First, although the contents of promises, orders and so on do not have truth values, they do have values: promises can be kept or not kept; orders obeyed or not obeyed, and so on. Using these values instead of truth values, if we get a table of values showing the same pattern as the truth-table for $p \supset q$, then I think we could say that the *if* in question corresponds to the hook. Alternatively, we could say that, if the propositions corresponding to the contents of the orders, promises, and so on have the same truth-table as $p \supset q$, then the *if* in question corresponds to the hook.

If we look at the two groups, we see that the *ifs* in group A are normal conditional *ifs* (they contrapose within the context of the speech act), whereas the *ifs* in group B are stipulative *ifs* (they do not contrapose within the context of the speech act). It might seem, in fact, that *ifs* occurring in speech acts of the type found in group A are always normal conditional *ifs* and those occurring in speech acts of the type found in group B are always stipulative *ifs*. This, however, is not the case. As we shall see, although normal conditional *ifs* occur only in acts with illocutionary forces such as the forces of the acts in group A and stipulative *ifs* occur only in acts with illocutionary forces such as the forces of the acts in group B, there are acts with the same kinds of illocutionary force as the acts in both groups A and B where the *if* in the content of the act is neither a conditional *if* nor a stipulative *if*.

In both groups A and B, it does seem that the *if* is not equivalent to material implication. In (1), in group A, my statement is true if, when I run, I pant; it is false if, when I run, I do not pant; but since I make no statement about what I do when I do not run, then when it is false that I run, my statement 'If I run I pant' is neither true nor false. Similarly, I predict what will happen if it rains; I make no prediction about what will happen if it does not rain. So my prediction comes true in the case when it does not rain and the match is cancelled and does not come true in the case when it rains and the match is not cancelled. When it does not rain, my prediction neither comes true nor does it not come true. Similarly with the guess. The same thing happens in group B. My promise is kept if, when you find my book, I give you a reward; it is not kept if, when you find my book, I do not give you a reward. In the cases where you do not find my book, the question of whether I have kept my promise or not does not arise. Similarly, with the order and the warning.

So, in the examples, in groups A and B, 'If p then q' and 'If p then $\sim q$' are contraries. Peter Downing (ANALYSIS 35.3) argues that this means that contraposition applies to *no* conditionals of whatever type. He gives J. L. Mackie's argument in *Truth, Probability and Paradox*: since 'If p then q' and 'If p then $\sim q$' are contraries, whereas 'If p then q' and 'If $\sim p$ then q' are compatible, if contraposition applies to

conditionals from this compatible pair we can deduce 'If $\sim q$ then $\sim p$' and 'If $\sim q$ then p', which are contraries. But, I would maintain, if 'If p then q' and 'If $\sim p$ then q' are compatible, they are not then conditionals. To maintain that 'If it rains the match will be cancelled' and 'If it does not rain, the match will be cancelled' are both true, is to deny that there is any connection between its raining and the match's being cancelled; it is equivalent to maintaining 'The match will be cancelled whether or not it rains'. In general, to say that 'If p then q' and 'If $\sim p$ then q' are both true is equivalent to saying 'q, whether or not p'. So I would still maintain that the *ifs* in group A contrapose.

There is another group of illocutionary acts where I think the *if* does correspond to material implication:

C.
(7) I state that if I did not go to Majorca in 1964 then I went to Greece
(8) I promise that if I do not give you a pony for your birthday I will take you for a trip to Greece
(9) I order you to peel the potatoes if you do not feed the hens

The contents of *all* of these acts can be contraposed in the context of the act. Furthermore, the statement I make in (7) is true not only in the case where it is true that I did not go to Majorca and true that I went to Greece, but also in the cases where it is false that I did not go to Majorca and true that I went to Greece and where it is false that I did not go to Majorca and false that I went to Greece. Similarly I keep the promise given in (8) not only in the case where I do not give you a pony and take you to Greece, but also in the cases where I give you a pony and take you to Greece and where I give you a pony and do not take you to Greece. The same is true of the order contained in (9). So in these examples in group C, we have *ifs* which are truth-functionally similar to material implication. Further, (7), (8) and (9) contain *ifs* which can be rendered in terms of *or*. Instead of (7), (8), (9), I could have said equally well

(7′) I state that I either went to Majorca in 1964 or I went to Greece
(8′) I promise either to give you a pony for your birthday or to take you for a trip to Greece
(9′) I order you to peel the potatoes or to feed the hens

In contrast, the *ifs* in the illocutionary acts in groups A and B cannot be rendered in terms of *or*. 'I state that either I do not run or I pant' is certainly not equivalent to 'I state that if I run I pant'; 'I promise that either you do not find my book or I will reward you' is certainly not equivalent to 'I promise to reward you if you find my book'.

The reason why there is this difference between the illocutionary

acts in group C and those in groups A and B would seem to be this. In group C two distinct states of affairs or two distinct acts are the objects of the illocutionary force-bearing verb, whereas in groups A and B only one state of affairs, or one act is the object of the illocutionary force-bearing verb, albeit a state of affairs or an act dependent on some condition, which condition is given in the dependent conditional clause. For example, in (7), the two states of affairs are my going to Majorca in 1964 and my going to Greece in 1964; in (4), in group B, the one act referred to is my rewarding you, but on the condition that you find my book.

Since '$p \supset q$' is defined as being equivalent to '$\sim p \vee q$', it is not surprising that, in illocutionary acts of the type in group C, the *if* in their contents can be interpreted as material implication. If the content of an act expresses a true disjunction, then 'if p then q' is equivalent to '$p \supset q$'.

The situation is this then. Illocutionary acts whose contents are propositions are the kinds of act in which we find normal conditional *ifs*; these are acts which have illocutionary force-bearing verbs such as 'I state', 'I predict', 'I guess' and so on. These are roughly the kinds of illocutionary force-bearing verb which are found in Austin's[1] classes of verdictives and expositives and also among his behabitives. Illocutionary acts whose contents are not propositions but which mention some act to be performed are the kinds of act in which we find stipulative *ifs*; these are illocutionary acts which have illocutionary force-bearing verbs such as 'I promise', 'I order', and so on. These are roughly the kinds of illocutionary force-bearing verb which are found in Austin's classes of exercitives and commissives and among the behabitives. But there are also illocutionary acts, containing either of the above types of illocutionary force-bearing verb, in which the *if* does correspond to material implication. These are acts whose contents can be rendered alternatively as disjunctions. We must notice here, however, that there are some *ifs* which can be rendered in terms of *or*, but an *or* which is not truly disjunctive. We find such *ifs* in threats, for example, '(I intend), if you do not give me your money, to take your life', which is more usually delivered in the form 'Your money or your life'. The *or* here is really *or else* and the *if* a stipulative one.

In the recent discussions of Ifs and Hooks in ANALYSIS, protagonists and antagonists have flung examples at each other, each claiming that his examples support his case. I do not want to discuss these examples here, but I believe that many of these supposed conflicts would be resolved if the authors would take into account the illocutionary forces of the examples they offer us.

[1] *How to Do Things with Words*, Oxford, 1962.

University of Nottingham

KIRK ON INDETERMINACY OF TRANSLATION

By M. C. BRADLEY

1. WHAT is the relation of underdetermination of physical theory to indeterminacy of translation? R. Kirk has recently sought to show[1] that the relation envisaged by Quine,[2] that namely between premiss and conclusion of a valid inference, does not obtain. In this paper I discuss Kirk's criticisms, and consider in passing some questions that arise about the interpretation of Quine.

2. In RIT Quine makes the indeterminacy turn on the underdetermination of physical theory by experience. At some point underdetermination sets in—though we are not required to agree on the place—and 'insofar as the truth of a physical theory is underdetermined by observables, the translation of the foreigner's physical theory is underdetermined by translation of his observation sentences. . . . Our translation of his observation sentences no more fixes our translation of his theory than our own possible observations fix our own physical theory' (RIT, 179). Kirk's reply is this. If indeterminacy must set in at the point at which underdetermination sets in, then however we are pleased to vary that point we will always have it that while everything below the point is translatable, everything above the point is subject to the indeterminacy.[3] To describe a case where the determined sentences are translatable but the underdetermined sentences also, is therefore to break Quine's argument from the fact of underdetermination to the existence of the indeterminacy. Such a case Kirk sets himself to describe. He supposes that we are translating Martian, and that underdetermination and therefore indeterminacy set in only with theoretical physics. Suppose now that we apply ourselves to translating their text-book.

> We could make a partial translation of the Martian text-book, as follows. We put into English everything except Martian theoretical words; and (as our assumption [viz. of translatability of all except theoretical terms] permits) we give the appropriate English syntax to those sentences in which theoretical and topic neutral expressions occur; but we leave the theoretical expressions in Martian. The result is a book differing from an English text-book only in that its theoretical vocabulary is unfamiliar.

[1] R. Kirk, 'Underdetermination of Theory and Indeterminacy of Translation', ANALYSIS 33 (1973) pp. 195–201. All page references in the body of the text prefixed by 'K'.

[2] W. V. Quine, 'On the Reasons for Indeterminacy of Translation', *Journal of Philosophy* 67 (1970) pp. 178–183. Hereafter referred to as 'RIT', with all page references in the body of the text.

[3] I imagine the sentences of a language arranged on a scale according to the likelihood of a philosopher's regarding them as underdetermined by experience. Presumably the observation sentences will be at the bottom and the logical truths at the top. Such an ordering is obviously only partial, and the programme necessarily vague, but should serve the present expository purpose, which is all that it is intended to do. A similar device is implied by Kirk.

> Now the theory M presented in this book might be isomorphic in the following sense with the English-speaking physicists' theory A: A theory presented in a foreign text-book, partially translated as described above, is in the relevant sense *isomorphic* with physical theory A if and only if the foreign and English theoretical vocabularies can be so mapped on to each other that, by word-for-word substitution according to this mapping, (i) The hitherto untranslated parts of the book yield sentences which fit into the context provided by those already translated in such a way that the result would be accepted by English-speaking physicists as a text-book of theory A; and (ii) The same goes, *mutatis mutandis*, for whatever supplementary explanations of the theory presented in the book the foreign physicists may be *disposed* to give.
>
> If the Martian theory M was in fact isomorphic with A in the sense explained, then the *only* difference between M and A would be one of vocabulary. But in that case M and A would be one and the same theory (except in a trivial sense). If mere difference of vocabulary constituted difference of theory, any fool could rival Einstein by substituting his own invented theoretical terms for Einstein's. Moreover, if mere difference of vocabulary were sufficient for theoretical difference, despite the sort of isomorphism described above, the conclusion of Quine's argument would be without philosophical interest or importance (K, 199).

3. Now the theory M in the *Gedankenexperiment* might indeed be isomorphic with A. The trouble is that nothing has been done to rule out the existence of other mappings, compatible with all speech dispositions but not isomorphic in Kirk's sense, and the problem remains of deciding between the various manuals of translation generated by these further mappings. A multiplicity of mappings is not ruled out by the existence of one mapping of a particular character, intuitively attractive though that particular mapping is.

4. While not acknowledging this objection Kirk does seek to avoid another difficulty. This is that 'isomorphism of text-books, or even of all actually uttered explanations of M and A would not of course ensure isomorphism of M and A in the sense defined, since dispositions might diverge' (K, 199). But, he replies, 'this normal inductive uncertainty must not be confused with Quinean indeterminacy. . . . By judicious questioning it would be possible for translators to become as clear about theory M as Martians are, and so to be justified in asserting its identity with A' (K, 199). But to claim this, he recognizes, is to run foul of Quine's view that 'no basis [of translation] can be gained by interrogation in a theoretical vein, since the interrogation would take place in the foreigner's language and so could itself be interpreted according to either plan [i.e. by either of two methods of translation]' (RIT, 180).

To this Kirk directs a series of replies which, bearing in part as they do on the objection I have proposed to him as well as on the one he has proposed to himself, must now be considered, though I adapt them a little to fit the changed context.

The first of these (K, 200) is that 'neither the indeterminacy thesis nor a special version of it covering the translation of physical theory may be included among the premisses of his [Quine's] argument'. The force of this is that what we are supposed to be reviewing is Quine's (one and only) argument *for* the indeterminacy thesis. This bears on my criticism of Kirk in an obvious way, since what that criticism assumes is that there *are* other manuals of translation of Martian physics all equally compatible with all Martian speech dispositions with respect to the sentences of physics. It assumes in particular that 'our interpretation of the explanations [given by interrogated Martians] will already be contaminated by our analytical hypotheses, and could not even in principle be corrected' (K, 200). Thus we may adduce against Kirk the possibility of other equally correct manuals of translation only by assuming the indeterminacy thesis. Yet it is the (one and only) argument to the indeterminacy thesis that we are supposed to be defending. Hence the proposed defence of Quine is guilty of *petitio*.

5. There are several lines of reply to this. One is that there is not, after all, just one argument to the indeterminacy, but at least *three*.[1] But a simpler reply is at hand. The charge was that we have met Kirk's objection to Quine's argument from underdetermination to indeterminacy by asserting that if Quine is right (in his thesis of indeterminacy) then Kirk's proposed counter-instance to that argument fails; more specifically, by asserting that if Quine is right then Kirk has not described a case where translation is determinate up to the point of underdetermination of physical theory, and determinate thereafter as well. But this is a slippery point; one might say that, on the contrary, Kirk is in *petitio*, since if Quine is right (about indeterminacy's being coterminous with underdetermination) we could reply on his behalf (as I have) that Kirk has not shown that, in his *Gedankenexperiment*, translation is determinate even beyond the point at which underdetermination of theory sets in, because he has not ruled out a multiplicity of translation manuals. Thus Kirk must reject the indeterminacy thesis in advance in order to structure his case as he needs to. The truth of the matter is that *petitio* is a poorly understood notion;[2] but it is hard to see how a correct theory of it would favour Kirk against Quine on the present point.[3]

[1] I discuss this question in 'Quine's Arguments for the Indeterminacy', forthcoming in *Australasian Journal of Philosophy*.

[2] The most strenuous attempt to unravel it that I know, by C. L. Hamblin in his book *Fallacies* (London: Methuen, 1970), seems to me seriously defective.

[3] I record, without further comment, a case comparable with the present one. J. Shaffer argued against Smart's identity thesis that we could have information about our own mental states without having information about our own cerebral states (*Journal of Philosophy* 58 (1961) pp. 813–822). Cornman replied in effect that this begged the question against Smart, since if Smart was right (in his thesis) then someone who has information about his mental states *eo ipso* has information about his brain states (*Journal of Philosophy* 59 (1962) pp. 486–492). Shaffer conceded the point (*Journal of Philosophy* 60 (1963) pp. 160–166).

6. The second of Kirk's replies to an objection (see §4 above) is that the view that interrogation in a theoretical vein is useless in deciding between manuals of translation holds only if 'it were also impossible for English-speaking pupils [by such interrogation] . . . to improve their chances of learning A rather than a different theory, B, from their instructors' (K, 200). Or, conversely, 'if English-speaking instructors can successfully teach their pupils A rather than B, then Martian ones can successfully teach ignorant translators their theory M rather than some other' (K, 200). Can English-speaking instructors successfully teach their pupils? Kirk gives two reasons for thinking that they can. If they cannot, then (i) 'nobody could possibly tell whether he or anyone else held theory A rather than theory B' (K, 200) and (ii) 'it would make no sense to say that these were different theories' (K, 200).

But it can hardly be supposed that (i) is a consequence that Quine would find unacceptable; where A and B are empirically equivalent theories the indeterminacy thesis, as applied domestically, holds precisely that the question whether a compatriot holds A rather than B is answerable only relative to a manual of translation. Moreover, in one's own case, Quine has argued at length that each of us knows what he means *himself* only relative to some background language.[1] These theses may indeed be untenable,[2] but that is for Kirk to show. (ii) fails also, since the distinctness of A and B does not have as a necessary condition that English teachers 'can successfully teach A rather than B'. The notion of *successfully teaching A rather than B* is anyway one that would fall into Quine's category of *the mentalistic*, since, as intended by Kirk, it amounts to something like *causing the learner to mean A by his words rather than B.*

7. I conclude that the reply I envisaged (§3 above) to Kirk's original objection to Quine survives the above two counters that are suggested in Kirk's paper (§4, §6 above). There is however another quite different point that might be urged against Kirk. His argument, as already observed, takes the form of supposing that underdetermination of physical theory sets in at a certain point, taking that as sufficient for translatability holding up to that point, then endeavouring to show that for that choice of point, translatability holds beyond it too. But it seems most doubtful whether this is after all a permissible form of argument against Quine, questions of detail aside. For one thing the problem of parsing the foreign language bears in that *wherever* underdetermination is supposed to set in, we still have some *initial* freedom of choice in the matter of 'what native devices to view as doing in their devious ways the work of our own various auxiliaries to objective reference: our articles and pro-

[1] W. V. Quine, 'Ontological Relativity', in *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969).

[2] I have argued in one way that the latter is in 'How Never to Know What You Mean', *Journal of Philosophy* 66 (1968) pp. 119–124.

nouns, our singular and plural, our copula, our identity predicate'[1] and since this freedom of choice apparently induces by itself a measure of indeterminacy (see footnote 1 on p. 20, *supra*), the total indeterminacy is not confinable, in Kirk's way, to the remote theoretical terms of the language. It seems that it will already be present at the bottom of the scale, regardless of the cut-off point for the determination of theory. But, anyway, we have only to look at the reasons given for the indeterminacy in the translation of quantification[2] to see that according to Quine, given indeterminacy of translation of terms *somewhere*, the translation of quantifiers is rendered indeterminate *everywhere*. Thus we cannot think with Kirk that translation can be freely supposed determinate up to any given point in the (theory) determination scale; given indeterminacy somewhere all quantified sentences *below* are made subject to indeterminacy also.

[1] W. V. Quine, *Word and Object* (Cambridge, Massachusetts: The M.I.T. Press, 1960), p. 53.
[2] Ibid., pp. 60–61.

University of Adelaide

CENTRAL-STATE MATERIALISM AND PARAPSYCHOLOGY

By John W. Godbey, Jr.

IT is sometimes argued that Central-State Materialism (CSM) would be falsified if the existence of parapsychological phenomena were established. For example, Keith Campbell has written that 'if even a single example of . . . paranormal phenomena is genuine, Central-State Materialism is false'.[1] Similarly, D. M. Armstrong gives several possible ways of 'explaining away' the data of parapsychology, and then remarks that 'if these ways of escape prove unsatisfactory, Central-State Materialism cannot be the whole truth about the mind'.[2] It is these assertions which I wish to refute in this paper.

CSM is one form of the Identity Theory of mind and body (IT). By the 'Identity Theory' I mean the theory which holds that all mental states are as a matter of fact identical with states of the body. Some philosophers, such as Smart and Armstrong, believe that the bodily states in question are states of the central nervous system, and this is why they call their materialism '*Central-State* Materialism'. Their theory

[1] *Mind and Body* (Garden City, New York, 1970), pp. 91–92. Page references to Campbell are to this book.
[2] *A Materialist Theory of the Mind* (London and New York, 1968), p. 364.

is a version of *materialism* because they believe that the central nervous system and all of its states and properties are physical or material, or physical neutral. (States or properties are 'physical neutral' if, even though they are not physical, non-living physical bodies can be in them, or have them. *Being beautiful* or *being the third member of a series* are examples of physical neutral properties.) The beliefs that mental states are identical with states of the central nervous system and that the central nervous system is solely physical or physical neutral are the heart of CSM. In order, then, to demonstrate the consistency of CSM and parapsychology I must (a) show that parapsychological phenomena do not imply the falsity of any of the identities asserted by the IT; and (b) show that these phenomena do not imply that the central nervous system is not solely physical.

(a) A number of parapsychological phenomena involve a person's allegedly coming to know something by a means other than reasoning, memory, sense perception or some combination of these. If the person comes to be aware of another person's mental states the phenomenon is called telepathy; if he comes to be aware of an event which will take place in the future it is called precognition; and so on. Let us suppose that precognition occurs: how could this refute the IT? Since the IT is only a claim that mental states are identical with bodily states, if precognition can refute the IT then from a description of a precognitive event it must follow that one or more of the identity statements are false. The one most relevant to precognition is

To know such and such is to be in brain state such and such.

But a statement asserting the existence of precognition is compatible with any statement which asserts that to know something is to be in a certain state. The existence of precognition could only show that one came to be in that state in a peculiar way. When we assert the existence of precognition we are asserting the existence or non-existence of certain causal relations; and since the IT neither asserts nor denies any causal claims, it cannot be falsified by the existence of precognition.

This argument can be generalized fairly easily. The vast majority of parapsychological phenomena can be placed in one of two mutually exclusive categories: the mind either comes to be in the state it is in (knowledge, belief, etc.) by paranormal means (e.g., precognition), or it causes something else to be in the state it is in by paranormal means (e.g., psychokinesis). But the acceptance of either of these kinds of phenomena, though it may cause a large revision or even rejection of parts of science or the philosophy of science, is compatible with the claim that the mental state in question is a brain state. Thus they are compatible with the IT.

(b) It follows that if parapsychological phenomena are incompatible

with CSM it can only be because the phenomena are incompatible with mental states being physical or physical neutral states—for this is all that CSM adds to the IT. And since all that the existence of the parapsychological phenomena would show is that certain causal chains do or do not occur, if Campbell's or Armstrong's arguments are sound they must show that physical or physical neutral states cannot enter into the appropriate causal chains. Neither of them—nor any other writer with whom I am familiar—has even attempted to show this. Rather, they argue that parapsychological phenomena are incompatible with modern science. Campbell's argument is typical:

> Parapsychological phenomena, by definition, demonstrate capacities of mind which exceed any capacities of brain. The brain is receptive only to information which arrives by neural pathways, and so is confined to perception by way of the senses. If some people can learn more about distant, hidden, or future fact than memory and inference from present sense perception can teach them, then their minds are not just brains. (P. 91.)

Campbell is mistaken. The fact that people can learn about distant facts other than by present sense perception, memory and inference does not show that their minds are not just brains. Such data would be consistent with, and in fact would seem to imply, merely that these people can acquire information in ways other than normal. Parapsychological data only demonstrate capacities of the mind which exceed any known capacities of the brain. Campbell acknowledges as much in his summary of his discussion:

> Even if some paranormal results were established as genuine, they might of course be accommodated in a new, expanded, physical science. . . . Television is paranormal with respect to Newton's physics, but not to ours.

But then he goes on:

> The fact that some neomaterialism might survive the establishment of paranormal truths would not vindicate Central-State Materialism. For Central-State Materialism is a materialism based on our present physical and chemical science. If that science is inadequate, the materialism based on it is false. (Pp. 96-7.)

What Campbell is apparently saying is that if present-day science is not completely true then CSM is false, because it is 'based' on present-day science. But in order to explain or define CSM no mention was made of science, or scientific theories, present-day or otherwise.

Campbell needed to show that if parapsychological phenomena occur, then the central nervous system cannot be material. The most he has shown is that if these phenomena occur, then present-day science cannot explain them. For his argument to have any force he would have

to show both that the phenomena could not be explained using present-day conceptions of the physical and that they could be explained under the hypothesis that humans are partly non-physical, or immaterial. Lacking a convincing argument for these two propositions, there is no reason to suppose that, even if they should be accepted as genuine, any of the phenomena under discussion could refute CSM. The likelihood of anyone producing arguments for these propositions appears to me to be remote. What could cause us to accept an unknown immaterial substance or property over an unknown physical one in explaining some phenomenon?

The parapsychological phenomena I have been discussing all, as I have said, fall into two classes: those phenomena in which the mind comes to be in some state by paranormal means, and those in which it causes something else to come to be in some state by paranormal means. Since I have shown that the existence of any or all of these phenomena is consistent with CSM, Campbell and those philosophers who agree with him are wrong in their belief that the existence of even a 'single example' of parapsychological phenomena would refute CSM.

I have not shown other possible kinds of parapsychological phenomena are compatible with CSM. However, until someone produces data to support a belief in the existence of another kind of phenomenon, and an argument to show its inconsistency with CSM, there is no reason to think that the results of parapsychological research are incompatible with CSM.

University of South Carolina

RULES OF LANGUAGE: REPLY TO MR. CARR

By C. H. WHITELEY

MR. C. R. CARR (ANALYSIS 35.2, pp. 51–3) is mistaken about the object of my article on Rules of Language (ANALYSIS 34.2, pp. 33–8). Whether or not there can be a 'general theory for finding out in a given case what a person's intentions are when his actions aren't sufficient to determine this' was not in my mind when I wrote the article, and I do not think I have written anything relevant to this question. Nor was I denying that there are 'sound-meaning pairings' which relate utterances to meanings in a systematic way. On the contrary, as Carr points out, my argument assumes that there are such rules. What I wished to maintain was that the skill of using a natural language, the skill of uttering meaningful sentences—that is, sentences that other people

can understand—and of understanding the meanings of sentences uttered by others, cannot be bounded by or summed up in any set of *strict* rules. For this skill includes the ability to utter sentences, and to understand sentences uttered by others, as having meanings not strictly derivable from any rules. If the purpose of a set of semantic rules is to describe what people do, or try to do, in using a language, any set of strict rules must misdescribe this. If its purpose is to instruct people in the skill of using the language, the rules cannot be rigid if they are to be effective.

Carr says that we must distinguish what a speaker means by uttering certain words from what the words themselves mean. There is a distinction to be made here, but it is not happily expressed. Words mean nothing of themselves, but have meaning only as they are intended and understood to convey some message. There is what Englishmen usually mean by 'corn' and what Americans usually mean by 'corn', but there is nothing that 'corn' itself means. When we distinguish between what a speaker meant by his words and what those words mean, we are distinguishing between what he meant to convey by the sounds he uttered, and what the majority of speakers would mean, and the majority of hearers understand, by sounds of the same type in most kinds of situation. A type-sentence has a usual meaning—or, more likely, a variety of usual meanings in a variety of situations; and these meanings are given by the semantic rules. But there may be some tokens of that type-sentence which are neither intended by the speaker nor understood by the hearer to have their usual meaning; and there is no justification for saying that *in this situation* 'the words themselves' mean something other than what they are intended and understood to mean. A Swiss acquaintance of mine, describing a man lying on a mountain path suffering from a heart attack, said that he was 'pumping'. What she intended to convey, and what I understood her to convey, was that he was gasping; and in this context this was what the word 'pumping' meant, notwithstanding that the semantic rules of English do not make 'gasping' and 'pumping' interchangeable. The distinction in this kind of case is not between what the speaker means and what the words mean, but between what the words would normally mean and what they mean on this occasion.

As to metaphors in general, an expression may have an established use in a language, and still be counted as metaphorical—for intance, to 'drive home' a point in an argument. It counts as metaphorical, not because it is a less common use than 'driving home a nail', nor because it is not covered by the rules of the language—the semantic rules of contemporary English must allow for this expression—but because it was at one time not covered by those rules, but was introduced as an analogical extension of the use relating to nails. If a man introduces a metaphorical expression which is original, and therefore not according to the semantic rules of the language he is using, one must say either that his

use is improper and the words do not have the meaning he gives to them, or that the rules of the language do allow him to make such metaphorical extensions, and are therefore not *strict*. I argue for the latter alternative. A readiness to understand such extempore extensions of meaning, and to make them when convenient, is part of proficiency in a language. The semantic rules do not strictly determine what can be intelligibly said in it.

University of Birmingham

QUINE'S WAY OUT

By PHILIP HUGLY *and* CHARLES SAYWARD

AS a way of dealing with the semantical paradoxes Quine suggests the following ([2], pp. 9–10):

> The expressions 'true', 'true of' and related ones can be used with numerical subscripts '0', '1', '2' and so on always attached or imagined. . . . Then we can avoid the antinomies by taking care, when a truth locution T is applied to a sentence S, that the subscript of T is higher than any subscript inside S. Violations of this restriction would be treated as meaningless, or ungrammatical, rather than as true or false sentences.

Despite the fact that this suggestion appeared some years ago we know of no in depth discussion of it in the literature. This is probably due to a failure to see the philosophical import of Quine's suggestion. In fact it has considerable philosophical import. This we establish in section I. Whether it is acceptable depends, among other things, on how two problems are dealt with. In section II we spell out one of these problems, and in section III suggest a way of dealing with it. We spell out the second problem in section IV. Right now it is unclear to us whether the second problem has a satisfactory solution.

I

Tarski regarded the semantical paradoxes as reason for thinking English and other comparably developed natural languages are inconsistent ([4], pp. 164–5). By calling a natural language inconsistent he meant that some subset of its true sentences entailed a sentence and its negation. His reasoning seems fairly expressed as follows: Consider these two principles:

(1) For every English sentence S obtainable from the formula '$Fx \longleftrightarrow y$' by replacing 'x' by a singular term designating a sentence, 'y' by that sentence: S is true.

(2) For every English sentence S obtainable from the formula '$xRy \longleftrightarrow z$' by replacing '$x$' by a singular term designating a unary predicate P, 'y' by a singular term b designating an object, 'z' by the result of attaching P to b: S is true.

If English contains any predicates satisfying (1) or (2), it is inconsistent. But the English predicates 'true' and 'true of' seem to satisfy (1) and (2). How then can we significantly theorize about English from a semantical point of view?

One tack is to admit that English, taken as a whole, is inconsistent, but to hold that various fragments of English, taken as separate wholes, are consistent and hence subject to significant semantical theorizing.

Thus, we might select two independently consistent fragments of English, designate one as object language and the other as metalanguage and construct within the latter a truth definition for the former, applying the methods and safeguards first laid down by Tarski. This would seem to be an approach sometimes favoured by Davidson ([1], p. 314).

Another approach, also suggested by Davidson ([1], p. 314), is to affirm the consistency of English by denying that any predicates formulable in English satisfy such semantical principles as (1) or (2). On this view we maintain a distinction among the sentences of English along the lines of truth and falsity, but deny to English the means of expressing and thus recognizing this distinction. The distinction would be drawn only in another language and the terms of that distinction would forever remain a portion of that other language untranslatable into English.

These approaches share the feature of "semantically reducing" the natural language—either by restricting one's theorizing to fragments of the language so chosen as to lack the means for semantical discourse about their own sentences, or by dealing with the whole language viewed as essentially lacking those means. The first approach goes against our feeling that we ought to be able to produce a single survey of the *whole* of what can be significantly said in a natural language. The second approach conflicts with the apparent fact that English (or any other comparably developed natural language) does provide its speakers with some significant purchase on the notions of truth and falsity in application to their own discourse.

Quine's suggestion stakes out a middle ground, and in this lies its philosophical import. We can view Quine as recommending that we consider, in place of English, a language like English except for certain small but crucial changes. We shall call this language revised English. One hope that motivates Quine's suggestion is that revised English will be consistent and hence open to a significant theory of truth of the sort set forth by Tarski. A second hope is that revised English will have the feature of universality remarked upon by Tarski ([4], p. 164) and reflected in our notion of the whole of what can be significantly said. Tarski thought that universality was both a characteristic feature of natural languages and the primary obstacle to giving them a significant truth theory.

Within revised English such principles as

(3) Let S be any sentence of revised English which results from 'x is true$_0$ $\longleftrightarrow y$' by replacing 'x' by a singular term designating a sentence and 'y' by that sentence: S is true$_1$

can be formulated. In effect revised English can speak about *any* of its semantical locutions, but cannot speak of all of its semantical locutions

for the cost of that would be contradiction. This can be seen from the fact that the principle just displayed must use 'true$_1$' and that other similar principles must use 'true$_2$', 'true$_3$' etc. To speak of *all* the one-place semantical locutions at once (rather than of any, one by one) would require *either* that revised English contain a non-subscripted locution 'true'—in which case the contradictions arise as in English—or that the subscripted numerals function as names of numbers and hence occupy a position open to quantification. But in that case too contradictions arise, as will be shown in section II.

Upon what basis can one hope that revised English has the feature of universality? First, if S is a sentence devoid of semantical locutions, then S belongs to English if and only if it belongs to revised English and thus has itself as its translation in revised English. Second, it appears that although revised English contains no sentences with unsubscripted semantical locutions, revised English will yet contain a translation of any such *non-paradoxical* sentence of English. Paradoxical sentences of English either get correlated with ungrammatical sequences of symbols or with false sentences of revised English. The upshot of all this is that it is plausible to think that revised English has the feature of universality if English does.

II

In part, Quine's suggestion amounts to proposing a revision of English which results in a language which has significant semantical resources and is yet open to a consistent theory of truth satisfying Tarski's Convention T. Were this condition not satisfied Quine's suggestion would lose much of its point. And a good case can be made for the claim that this condition is not satisfied. The case is based on these premisses:

Premiss 1. No Tarski style theory of truth can be given for a language with infinitely many semantically primitive predicates.

Premiss 2. A language revised in accord with Quine's suggestion has infinitely many semantically primitive predicates.

Premiss 1 seems trivially true, for a semantically primitive predicate P is one such that a truth definition for a language L less P does not suffice for L. Hence P will need its own clause in a definition of 'satisfaction' adequate for L. Thus, if L has an infinite number of such semantically primitive predicates, the component of the truth definition specifying satisfaction would need infinitely many clauses and thus could never be completed. (A semantically primitive predicate need not be a semantical predicate, though in the case at hand the idea is that the sub-

scripted semantical predicates suggested by Quine will each be semantically primitive.)

Consider now premiss 2. For the sake of simplicity focus on the following rules which result from applying Quine's suggestion to 'true'.

(*i*) 'True_0' is a truth locution of revised English.

(*ii*) For any numeral n, if $\ulcorner \text{true}_n \urcorner$ is a truth locution of revised English so is $\ulcorner \text{true}_{n+1} \urcorner$.

(*iii*) When a truth locution T is applied to a sentence S of revised English, the result is ungrammatical unless the subscript on T is higher than any subscript inside S.

Rules (*i*) and (*ii*) together generate infinitely many truth predicates. But there is nothing in any of the rules, nor in anything else Quine says with respect to his suggestion, which tells us how to get from the meaning of $\ulcorner \text{true}_m \urcorner$ to $\ulcorner \text{true}_n \urcorner$ for distinct numerals m and n. We are thus left with an infinity of expressions each of which, without explanation to the contary, should be regarded as semantically primitive. Accepting Quine's suggestion must await that explanation.

In effect our suggestion is that the numerals, in their use as subscripts to 'true', 'true of' etc., function merely as distinctive shapes and hence as alphabetic elements. Rules (*i*) and (*ii*) are essentially rules for spelling. Rule (*iii*) limits the possible combinations of predicates with expressions in terms of certain alphabetic features of the predicates and expressions. So regarded, it is clear that we are prima facie dealing with infinitely many semantically primitive predicates. The explanation to the contrary is really needed.

The supposition that the numerals are not being used merely as providing an infinite stock of distinctive shapes, alphabetic elements, would, on the other hand, seem to come to this: that in subscripted position they are still functioning in a denotative capacity, namely as names of numbers.

To help bring out the initial plausibility of this idea we can think of revised English as dividing up into a series of ever more encompassing language levels. An expression of level 0 would be any expression devoid of subscripted semantical locutions. An expression of level 1 would be any expression of level 0 together with all grammatically permissible combinations of level 0 expressions with semantical locutions using the subscript '0'. And so forth. On this view

$(n)(x)(x$ is true_n (in revised English) $\longleftrightarrow$ x is a true sentence of level n (of revised English))

would be the case. Thus, on this view the subscripted numerals are not mere alphabetic elements. They would be meaningful expressions, names of numbers.

But could this be correct? One point which should rouse our suspicions is that on this account it seems we could define 'true' *simpliciter*:

x is true $\longleftrightarrow$ $(\exists n)$ (x is true$_n$).

That is, we could say that a sentence x is true just in case there is some level n of the language such that x is a true sentence of that level. It seems we could similarly define 'true of' *simpliciter*:

x is true of y $\longleftrightarrow$ $(\exists n)$ (x is true$_n$ of y).

The question, then, is whether, e.g., we should regard '$(\exists n)$ (x is true$_n$)' as a genuine truth predicate of revised English. Consider now the following metalinguistic argument. Suppose that y is a true sentence of revised English. Then, for some number N, y is a true sentence of level N. Hence, where x is a singular term designating y and n is a numeral designating N, $\ulcorner y \rightarrow (\exists n)$ (x is true$_n$)$\urcorner$ is a true sentence of revised English. Next suppose that $\ulcorner(\exists n)$ (x is true$_n$)$\urcorner$ is a true sentence of revised English. Then y is a true sentence of level N. Hence $\ulcorner(\exists n)$ (x is true$_n$) $\rightarrow y\urcorner$ is a true sentence of revised English; and so, consequently, is $\ulcorner y \longleftrightarrow (\exists n)$ (x is true$_n$)$\urcorner$.

A similar argument yields a parallel conclusion for '$(\exists n)$ (x is true$_n$ of y)'. We can thus introduce these two metalinguistic principles for revised English.

(4) For every sentence S of revised English which results from '$(\exists n)$ (x is true$_n$) $\longleftrightarrow$ y' by replacing 'x' by a singular term designating a sentence and 'y' by that sentence: S is true.

(5) For every sentence S of revised English which results from '$(\exists n)$ (x is true$_n$ of y)' by replacing 'x' by a singular term designating a unary predicate P, 'y' by a singular term b designating an object: (a) If $\ulcorner Pb\urcorner$ is grammatical, $\ulcorner Pb \longleftrightarrow S\urcorner$ is true; (b) if $\ulcorner Pb\urcorner$ is not grammatical, S is false.

But both (4) and (5) are paradoxical principles. This is most easily shown in connection with (5). For both 'x' and 'y' we substitute the singular name '$\ulcorner\sim(\exists n)$ (x is true$_n$ of x)$\urcorner$'. For these substitutions clause (a) of (5) applies. Thus, we have as true by (5) the following sentence,

$(\exists n)$ ('$\ulcorner\sim(\exists n)$ (x is true$_n$ of x)$\urcorner$' is true$_n$ of '$\ulcorner\sim(\exists n)$ (x is true$_n$ of x)$\urcorner$') $\longleftrightarrow$ $\sim(\exists n)$ ('$\ulcorner\sim(\exists n)$ (x is true$_n$ of x)$\urcorner$' is true$_n$ of '$\ulcorner\sim(\exists n)$ (x is true$_n$ of x)$\urcorner$'),

which is a contradiction.

To stave off inconsistency, then, we must regard the subscripted numerals as mere alphabetic elements, themselves devoid of denotation. And in that case rules (*i*) and (*ii*) *are* mere spelling rules for writing

infinitely many predicate symbols, and this surely gives force to the suspicion that these predicates may one and all be semantically primitive.

III

In this section we provide a basis for allaying the doubts just raised. Borrowing some ideas from Raymond Smullyan in [3] we describe a very simple language S with just enough structure to be inconsistent for reasons provided by the liar paradox. We revise S in accordance with Quine's suggestion. We then show that, although revised S has infinitely many truth predicates, a single clause of the truth definition for this language settles the truth conditions for all sentences into which they enter. This shows that premiss 2 is false.

The basic idea we borrow from Smullyan is what he labels the 'norm function'. The norm of an expression is the expression followed by its own quotation. We want S to contain an analogue of

> 'Does not yield a truth when applied to its own quotation' does not yield a truth when applied to its own quotation.

Thus S has to contain negation, its own truth predicate, the means of designating each of its own expressions and, finally, the means by which some of its expressions are self-designating. It is in connection with the latter requirement that the norm function comes in.

Language S

Syntax

1. E is an *expression* iff E is a finite sequence of the signs: $P, T, N, *, \sim$.
2. Q is a *quotation* of E iff $Q = \ulcorner *E* \urcorner$.
3. n is a *norm* of E iff $n = \ulcorner E*E* \urcorner$.
4. d is a *designator* iff d is a quotation or a quotation preceded by one or more occurrences of 'N'.
5. X is a *sentence* iff X = 'P' or d is a designator and $X = \ulcorner Td \urcorner$, or Φ is a sentence and $X = \ulcorner \sim\Phi \urcorner$.

Semantics

6. The quotation of E ($=\ulcorner *E* \urcorner$) designates E.
7. If an expression E designates an expression F, then $\ulcorner NE \urcorner$ designates the norm of F ($=\ulcorner F*F* \urcorner$).
8. 'P' is true iff $2 + 2 = 4$.
9. $\ulcorner Td \urcorner$ is true iff the object designated by d is true.
10. $\ulcorner \sim\Phi \urcorner$ is true iff Φ is not true.

On the basis of this semantics, 'P' is a true sentence of S. '$T*P*$' is also a true sentence of S since '$*P*$' is a quotation which designates 'P' which is a true sentence of S. However, by application of several clauses above we can derive the following biconditional:

'$\sim TN^*\sim TN^*$' is true iff the norm of '$\sim TN$' is not true.

Thus, since the norm of '$\sim TN$' = '$\sim TN^*\sim TN^*$' we can conclude:

'$\sim TN^*\sim TN^*$' is true iff '$\sim TN^*\sim TN^*$' is not true.

which is a contradiction.

Thus S is shown to include among its true sentences a contradiction. S is inconsistent. We attack this problem by revising S along the lines of Quine's suggestion. We add a new basic sign: "'". We then define the *level of a truth predicate* for revised S ('L' for short)

'T' is a truth predicate of level 0,
If X is a truth predicate of level n, $\ulcorner X' \urcorner$ is a truth predicate of level $n + 1$,

and the *level of an expression*

E is of level 0 iff 'T' does not occur in E,
E is of level n iff every truth predicate in E is of level $n - 1$ or less.

Thus, a given level n of L includes all expressions of all lower levels and excludes only those expressions containing truth predicates of level n or higher. We next rewrite a clause in the definition of 'sentence',

If d is a designator and X is a truth predicate, then $\ulcorner Xd \urcorner$ is a sentence iff the level of d is less than or equal to the level of X.

With this revision '$\sim TN^*\sim TN^*$' is excluded from the class of sentences. We now reformulate the crucial semantical clause in the following obvious way:

If X is a truth predicate of a level greater than that of a designator d, then $\ulcorner Xd \urcorner$ is true iff the object designated by d is true.

The only sentences in L are 'P' and the results of appropriately attaching truth predicates to designators of expressions of L and the negations of such expressions. Sentences of level 0 are only 'P' and the subsequent negations. For each of these the truth definition for L provides truth conditions. The sentences of level 1 other than those of level 0 all consist of attaching 'T' to designators of level 0 together with the negations. The negations are given truth conditions if the non-negations are. A non-negation is true if and only if its designator designates 'P' or one of the double negations of 'P'. So all these cases are also handled. In short, the truth theory for revised S is formally correct and materially adequate.

A stronger conclusion than the one we have established is this: If a formally correct and materially adequate theory of truth can be given for a language without its semantical locutions, then such a theory can be given for that language plus semantical locutions stratified in accord with

Quine's suggestion. But our investigation of revised S does show that a language's having the means of designating each of its own expressions plus being a language in which self-reference is possible are not enough to make the difference. And this certainly has to constitute evidence for the general thesis. Whether there are other features of a language which would make the difference is an open question. We consider a possibility next.

IV

Crucial to Quine's proposal is the rule that disallows certain applications of the truth locutions:

(*iii*) When a truth locution T is applied to a sentence S of revised English the result is ungrammatical unless the subscript on T is higher than any inside S.

As it stands the rule is not altogether clear. What is to count as an application of a truth locution to a sentence? Consider

(6) ' "Snow is white" is true_0' is true_0

which is ruled ungrammatical by (*iii*). What are we to say of

(7) The third sentence John spoke to Mary is true_0,

supposing

(8) The third sentence John spoke to Mary = ' "Snow is white" is true_0'

is true? Does (7) count as an application of 'true_0' to ' "Snow is white" is true_0'?

Suppose the answer is no. Then an unambiguous rendering of (*iii*) would be

(*iv*) For any numeral n, if S is a sentence of revised English and Q is the quotation name of S, $\ulcorner Q \text{ is true}_n \urcorner$ is ungrammatical unless any subscript m occurring in S is such that $\ulcorner m < n \urcorner$ is true.

This rule does not apply to (7). However a theory of truth for revised English must contain some rule that does apply to (7). And this further rule must rule (7) ungrammatical. Otherwise 'true_0' would be a kind of intensional verb: we could move from sense to nonsense by interchange of coreferential singular terms.[1] It is clear to us that Quine would not

[1] The same difficulty would arise were revised English further revised so as to eliminate singular terms in accordance with Quine's thesis about the eliminability of singular terms. We have given the argument in terms of the interchange of singular terms for two reasons: (*a*) Quine does not avail himself in the passage under investigation of the eliminability thesis, and (*b*) to give the argument relative to a notation free of singular terms would make our discussion more technical without increasing its perspicuity.

accept this result. Accordingly he must accept the further rule which in conjunction with (*iv*) yields:

(*v*) For any singular term *b* of revised English and for any numeral *n*, ⌜*b* is true$_n$⌝ is ungrammatical in revised English if *b* denotes a sentence with a numeral *m* such that ⌜*m* > *n*⌝ is true.

Suppose, on the other hand, (7) *does* result from applying 'true$_0$' to '"Snow is white" is true$_0$'. Then it is just (*v*) which is an unambiguous expression of (*iii*). So in either case Quine seems forced to accept (*v*). And herein lies the problem: for a number of reasons we think Quine would not want to accept (*v*).

First, it is anomalous to have a rule of purely syntactic import utilize a semantical concept. The effect would be, for example, that to know whether 'The third sentence John spoke to Mary is true$_0$' is *grammatical* we should first have to determine a matter of extra-linguistic fact, namely, which sentence was the third sentence John spoke to Mary. To know whether one sentence made sense, was grammatical, we would have to know whether another sentence was true.

Second, and more serious, it is unclear how (*v*) could occur in a truth theory for revised English. On the one hand, it seems as if it should be part of the clause defining sentencehood for revised English. But its entry would first require introduction in the metatheory of the singular term 'denotes'. But this seems impossible since the semantical component of the theory presupposes prior completion of the syntactical component.

Finally, since 'denotes' is a semantical locution its occurrences in revised English must also be subject to subscribing rules. This will raise problems of circularity when we attempt to formulate (*v*) in revised English itself.

Consider the following four properties:

Property 1: If a correct theory of truth can be given for language *L* devoid of semantical locutions, such a theory can also be given for *L* plus its semantical locutions stratified.

Property 2: For any expression *E* of *L* there exists a singular term in *L* designating *E*.

Property 3: Some expressions in *L* designate themselves.

Property 4: For any expression *E* of *L* there exist indefinitely many singular terms in *L* designating *E*.

One big difference between English and revised *S* is that English has property 4 and revised *S* does not. In the previous section it was argued that a language's having properties 2 and 3 does not preclude its having

property 1. Now it seems that a language's having property 4 may preclude it from having property 1. This depends upon how serious are the three difficulties with (*v*), especially the second one. Determining this decisively has to await further work. We hope we have said enough here to make such research seem worthwhile.

REFERENCES

[1] Donald Davidson, 'Truth and Meaning', *Synthese* 17 (1967), 304–323.
[2] W. V. Quine, 'The Ways of Paradox', *The Ways of Paradox*, Random House, New York, 1966, 3–21.
[3] Raymond Smullyan, 'Languages in which Self-Reference is Possible', *Journal of Symbolic Logic*, 22 (1957), 55–67.
[4] Alfred Tarski, 'The Concept of Truth in Formalized Languages', *Logic, Semantics, Metamathematics*, Oxford University Press, 1956, 152–278.

University of
Nebraska—Lincoln

A NOTE ON WITTGENSTEIN AND CYCLICAL COMPARATIVES

By DENNIS JAY PACKARD

IN *Philosophical Investigations* Part I, Wittgenstein writes

> Instead of producing something common to all that we call language, I am saying that these phenomena have no one thing in common which makes us use the same word for all—but that they are related to one another in many different ways (§ 65).

An illustration of this idea given by Bambrough in his paper 'Universals and Family Resemblances' (*Proceedings of the Aristotelian Society*, 1960–61) is as follows: A botanical taxonomist could classify a set of objects by reference to the presence or absence of features A_1, A_2, A_3, A_4, A_5. It could well happen that each of five objects, *a, b, c, d, e,* has four of these features and lacks the fifth, and that the missing feature is different in each of the five cases, and yet they are all classified under a single classification, *P*. That is, we could have the following situation:

P	*a*	*b*	*c*	*d*	*e*
	A_1	A_2	A_3	A_4	A_5
	A_2	A_3	A_4	A_5	A_1
	A_3	A_4	A_5	A_1	A_2
	A_4	A_5	A_1	A_2	A_3

So A_1 through A_4 are possessed by a, A_2 through A_5 are possessed by b, and so on. But as the following chart shows, none of the features A is possessed by all five objects

P	A_1	A_2	A_3	A_4	A_5
	c	d	e	a	b
	d	e	a	b	c
	e	a	b	c	d
	a	b	c	d	e

That is, A_1 is possessed by c, d, e, a but not b and A_2 is possessed by d, e, a, b but not c, and so on. Bambrough concludes: 'Here we can already see how natural and how proper it might be to apply the same word to a number of objects between which there is no common feature.'

Bambrough's example illustrates Wittgenstein's idea of family resemblance for properties. The question that naturally arises is: what about family resemblance for relations? One way of viewing this is to consider relations as properties of ordered pairs of objects. For example x being preferred to y can be viewed as a property P of preference holding for the ordered pair (x, y). With this in mind, consider the following Bambrough-type example:

Let each A in the previous example be a relation R_i. Let the objects a, b, c, d, e be the pairs of objects (v, w), (w, x), (x, y), (y, z), (z, v), respectively. Let P be a relation. Using the same configuration as before, we get the following:[1]

P	(v, w)	(w, x)	(x, y)	(y, z)	(z, v)
	R_1	R_2	R_3	R_4	R_5
	R_2	R_3	R_4	R_5	R_1
	R_3	R_4	R_5	R_1	R_2
	R_4	R_5	R_1	R_2	R_3
	R_1	R_2	R_3	R_4	R_5
	(x, y)	(v, z)	(z, v)	(v, w)	(w, x)
	(y, z)	(z, v)	(v, w)	(w, x)	(x, y)
	(z, v)	(v, w)	(w, x)	(x, y)	(y, z)
	(v, w)	(w, x)	(x, y)	(y, z)	(z, v)

(Just as an example consider these objects to be cars, the R_i to be factors such as being better built, riding more smoothly, being more beautiful, etc., and xPy to mean that x is a better car than y; or, if one liked to stay with family resemblance examples, let xPy mean x is a better game than y and let the R_i be factors used in evaluating games.)

[1] This pattern is just the voting paradox for five voters.

Just as before, P is true of each of these ordered pairs because of their mutual similarities; that is, vPw, wPx, xPy, yPz and zPv. The conclusion is: just as before, these pairs have no relation R_i that holds of them in common. But further, the relation P admits a cycle. Thus, even if we were to assume that the relations R_i were acyclical, there would be none of these relations had in common by these ordered pairs to ensure that P would be acyclical. Our "craving"[1] for acyclicity in P could be because we feel that these ordered pairs must have in common some other relation than P and that this relation must be something basic like length, weight, size, etc., which, being an acyclical relation, would force P to be acyclical. If, however, we give up the idea of requiring there to be some other relation common to the ordered pairs, the way is left open for cycles.

Cyclical comparatives were first discussed by political theorists concerned with voting procedures.[2] In fact, considering the R_i of the last example to be the votes of citizens 1 to 5 on options x, y, z, v and w, and considering P to be society's preference on the basis of voting, such an example is known in the literature as the *voting paradox*. That is, since x is ranked higher than y by citizens 1, 3, 4, 5, x is preferred by society on the basis of voting, i.e., xPy. Similarly, yPz, zPv, vPw and wPx. Thus we have a cycle. In the 50's, it was May who first pointed out that such examples can also be viewed as an individual basing his judgment on a number of factors, e.g., on the factors suggested for evaluating goodness of cars.[3] Recently psychologists have constructed experimental situations in which subjects consistently expressed cyclical preferences. Such considerations tend to suggest that cyclical comparatives are not unreasonable.[4] But if they aren't, the question arises as to why they should seem paradoxical. To be sure, part of the feeling of paradox is generated by the fact that for such comparatives the usual explications of the superlative are unacceptable. That is, using Quine's notations, let F_{est} in A be the best options in A or the most preferred options in A, most beautiful paintings in A and so on according to whatever comparative is under consideration.[5] Let F_{er} be the comparative under consideration. Now,

[1] In the Blue Book, pp. 17–18, Wittgenstein writes: 'This craving for generality is the resultant of a number of tendencies connected with particular philosophical confusions. There is—(a) The tendency to look for something in common to all the entities which we commonly subsume under a general term.... (b) There is a tendency rooted in our usual forms of expression, to think that the man who has learned to understand a general term, say, the term "leaf", has thereby come to possess a kind of general picture of a leaf, as opposed to pictures of particular leaves.'

[2] Condorcet, Marquis de, *Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix*, Paris, 1785.

[3] May, K. O., 'Intransitivity, Utility, and the Aggregation of Preference Patterns', *Econometrica*, XXII, 1954.

[4] Coombs, Clyde H., Dawes, Robyn M., Tversky, Amos, *Mathematical Psychology: An Elementary Introduction*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1970, pp. 159–163.

[5] Quine, W. V., *Philosophy of Logic*, Elizabeth and Monroe Beardsley, eds., Prentice-Hall Foundations of Philosophy Series, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1970, p. 78.

whatever F_{est} in A is, it is sure to be a subset of A. It might in fact be A, or one might be able to narrow it down. But surely, if A isn't empty, F_{est} in A won't be narrowed down to the empty set; that is, if $A \neq \phi$, then F_{est} in $A \neq \phi$. Here is a typical explication of the superlative

$$x \in F_{est} \text{ in } A \text{ iff } x \in A \text{ and not } \exists y \in A\ (F_{er}(y, x)).$$

But let A be the set of elements v, w, x, y, z in the last example and let F_{er} be the relation P in that example. Then it follows that since there is a cycle, $\forall x \in A\ \exists y \in A F_{er}(y, x)$. So according to the above explicaton F_{est} in $A = \phi$. But this can't be, so the explication is unacceptable. There have been proposals, however, in the literature as to how this explication can be suitably modified to handle such cases.[1]

The idea of family resemblance suggests another reason for the feeling of paradox about cyclical comparatives, the reason alluded to previously. Just as one might feel that being a game or not is based on a set of basic factors and that some of these basic factors must be had in common by all games, so one might feel that x being a better game than y is based on a number of basic factors some of which must be had in common by all ordered pairs (x, y) such that x is a better game than y. Such basic factors would be likely to include basic comparatives; since basic comparatives, e.g., length, weight, etc., tend to be acyclical, there must be some such basic transitive comparative had in common by all such pairs. Hence, the relation of being a better game (or of being preferred, etc.) must be acyclical. So the line of reasoning goes. But if cyclical comparatives are not unreasonable, then something must be wrong with such an argument. A Wittgensteinian answer might be that an error was committed in assuming that ordered pairs that fall under a particular comparative need necessarily have anything more in common than family resemblance.

[1] For alternative explications of this concept and for further discussion of the issues involved see: Schwartz, Thomas, 'Rationality and the Myth of the Maximum', *Noûs*, May, 1972; Packard, Dennis, 'Rational Ranking Functions for Cyclical Comparatives', Dissertation, Stanford, 1974; id. 'Social Choice Theory and Citizens' Intransitive Weak Preference—A Paradox', to appear in *Public Choice*.

Brigham Young University

EXPLANATION AND CONFIRMATION AGAIN

By MICHAEL MARTIN

IN a recent study of confirmation, Brody[1] suggested the following two conditions

(1) If evidence E confirms H_1 and H_1 explains H_2, then E confirms H_2.

(2) If evidence E confirms H_1 and H_2 explains H_1, then E confirms H_2.

Brody intended (1) and (2) to replace the following two confirmation conditions which resulted in certain paradoxes

(1′) If evidence E confirms H_1 and H_1 entails H_2, then E confirms H_2.

(2′) If evidence E confirms H_1 and H_2 entails H_1, then E confirms H_2.

In ANALYSIS 32.5 I criticized (1) and (2) and Barry Gower has now (in ANALYSIS 33.3) criticized my critique. Gower says that I argue that there are two sorts of cases that show that (2) is false: (a) cases where E confirms H_1 and H_2 explains H_1, but E disconfirms H_2, and (b) cases where E confirms H_1 and H_2, but H_2 does not explain H_1. Gower dismisses cases of type (b) on the grounds of irrelevancy 'since condition (2) only specifies conditions which are sufficient for confirmation' (p. 107). He argues that I have failed to give a genuine case of (a).

Gower's dismissal of cases of type (b) can be handled easily. I did not intend cases of this sort to show that (2) is false. Rather, I used them explicitly (p. 169) to tell against Brody's claim that (2) allows 'all inferences for which the converse consequence condition was used and needed'. I cited a case which on intuitive grounds one wants to allow, a case in which E confirms H_1 and H_2 entails H_1 (but does not explain H_1) and in which E also confirms H_2. This is not allowed by (2), but it is allowed by (2′), the converse consequence condition.

Gower argues against my cases of type (a) on two grounds. First, according to Gower, my example presupposes (1′) which Brody has replaced by (1). Gower argues that consequently I have not shown that (2) is false but only that (2) is incompatible with (1′). Secondly, he argues that even if (1′) is allowed my example does not show that H_2 explains H_1, since on my own showing E_1 disconfirms H_2. But H_2 is problematic as an explanation of H_1 if H_2 is disconfirmed.

[1] B. A. Brody, 'Confirmation and Explanation', *Journal of Philosophy*, 1968, pp. 282–299.

However, Gower is mistaken to suppose that I have not shown that (2) is false and that all I have shown is that (2) is incompatible with (1'). If my example is plausible then it does show that (2) is false, whatever else it might show. Perhaps what Gower means is that in order to show that (2) is false by using my example one must assume that (1') is true. But this is not quite correct either. What I must assume is only that there are certain inferences that are correct and that are allowed by (1'). I need not assume that (1') is true.

On the second point it is not obvious to me that if H_2 is disconfirmed H_2 cannot explain H_1. Hempel,[1] in some of his most famous writing on explanation, requires that the premisses of an explanation be true, not well confirmed. Truth is compatible with disconfirmatory evidence. Whether Brody intended 'explain' to be understood in this objective sense, the sense in which one thing explains another independently of the evidence available to an observer, is unclear. If he did not, it is up to him or his supporter to tell us how 'explain' is to be understood. Moreover, even if high confirmation is required of explanatory premisses, some negative evidence is compatible with high confirmation. Consider the example I used

> E_1 Mr. X lives in Chinatown and most people who live in Chinatown are law-abiding and most people who live in Chinatown are not fundamentalist Baptists.

E_1 disconfirms

> H_2 Mr. X is a fundamentalist Baptist and most fundamentalist Baptists are law-abiding.

However, the confirmation of H_2, at least for explanatory purposes, presumably should rely on the *total* available evidence, of which E_1 may be a very small part.

Consider another body of evidence E_5 that is independent of E_1 and makes (the relevant part of) H_2 virtually certain relative to this evidence. For example, there may be thousands of eye-witnesses to Mr. X's regularly attending a fundamentalist Baptist church. Now E_1 and E_5 when combined may still make H_2 likely. The existence of E_5 is quite compatible with E_1, and nothing I said in my critique rules E_5 out. For some unknown reason Gower thinks that E_1 is the only relevant evidence for H_2. In fact it is simply the only relevant evidence mentioned in the example.

[1] C. G. Hempel and Paul Oppenheim, 'Studies in the Logic of Confirmation', *Readings in the Philosophy of Science*, ed. H. Feigl and M. Brodbeck, p. 322.

Boston University

ON A PROPOSED REFUTATION OF HUME

By SARAH WATERLOW

D. C. STOVE has published a formal argument,[1] constructed originally by Manfred von Thun, which, it is claimed, refutes the 'inductive scepticism' of Hume. By 'inductive scepticism' Stove means the view that predictions based upon experience are no more reasonable, i.e. no better grounded, than predictions, or rather guesses, based on no experience at all. 'Inductive irrationalism' is perhaps a better name for this than 'inductive scepticism'. For the view implies that we have no rational ground for relying on induction as a source of knowledge, not merely that we may not have any ground or cannot know that we do. It is not my purpose here to enter into a full discussion of Hume as inductive irrationalist. Nor will I touch at all upon the wider question of the rationality of induction itself. I assume throughout that Hume does hold a position appropriately entitled 'inductive irrationalism', and I shall be solely concerned with Stove's characterization of this position and the consequences to which that characterization gives rise. The discovery of a formal refutation of Hume's inductive irrationalism would be an important discovery indeed. But it will here be argued that the von Thun-Stove refutation fails through a misunderstanding of the position which it is intended to disprove.

The proposed refutation is based entirely on propositions of logical probability which Stove maintains are intuitively obvious *a priori*.[2] These include (a) *principles* of logical probability such as the Negation Principle '$P(q, p) = 1 - P(\sim q, p)$' (i.e. 'The argument from "p" to "q" is less than wholly conclusive by the same degree as that by which the argument from "p" to "$\sim q$" is conclusive'); and (b) *statements* of logical probability which can be known to be true *a priori*, as for instance that $P(s, t) < 1$ where 's' is contingent and 't' a tautology.

These propositions are obvious enough by any ordinary standard: but is Stove entitled to take them for granted as premisses to an argument against *Hume*? For instance, they have meaning as propositions of logical *probability* only so long as it is allowed that the ascription of different possible degrees of conclusiveness to logically invalid arguments makes sense.[3] Yet does not Hume's inductive irrationalism imply that from a strictly rational standpoint this makes no sense at all? If so, any argument directed against the latter position would be flawed at start by *petitio principii* so far as it relied on propositions such as those just mentioned.

[1] *Probability and Hume's Inductive Scepticism*, by D. C. Stove, Oxford 1973, pp. 68–73.
[2] Ibid., pp. 66–67; cf. pp. 18–19.
[3] Compare Stove's remarks on p. 9.

However, the von Thun-Stove refutation is immune to this particular objection. Stove has carefully delineated Hume's position on induction in such a way as to show that there is no need to take it as implying that invalid arguments cannot vary in degree of conclusiveness, or that the concept of varying degrees of conclusiveness lacks sense. Rather, Hume's position is, essentially, the weaker assertion that no invalid empirical argument is rendered more conclusive by addition to the premisses of any data whatsoever, so long as the expanded premisses still fail to entail the conclusion.[1] Hume, on this view, could consistently assent to a proposition such as '$P(F(a)\ \&\ G(a),\ H(b)) < P(F(a),\ H(b))$' (where the component propositions are contingent and mutually independent).[2] Hume could therefore consistently agree that in some cases a logically invalid argument may be rendered more conclusive, though still not valid, by withdrawing an independent conjunct from the conclusion in cases where the conclusion contains independent conjuncts. But to strengthen an argument in this way is not to strengthen it *inductively*. Faith in induction means faith that arguments can be strengthened by widening the premisses to include further data. It is precisely this possibility that Hume, in Stove's view, is denying. Since the denial does not entail rejection of the concept of differing degrees of conclusiveness as such, Stove is perhaps not unjustified in employing propositions of logical probability in his attack on Hume's inductive irrationalism. (Whether such propositions, or an argument based on them, would survive scrutiny in the light of Humean *de*ductive irrationalism, is another question.)

At any rate, I shall assume that Stove is entitled to the premisses of his refutation. Nor is the validity of his reasoning in doubt. But the refutation, I shall argue, is misdirected, at least if intended to be directed at Hume.

Stove gives the following account of Hume's position: Hume holds that (1): If we have never experienced conjunctions of heat and flame, then when a single instance of flame is presented, the inference to the presence of heat as yet unexperienced is no more conclusive than an inference to the presence of heat from no empirical data at all; and (2): Given experience of past conjunctions (Flame-Heat), we do in fact infer to the presence of heat from a presented instance of flame; but the inference is no more conclusive than an inference from a presented instance of flame alone. 'Now it seems evident that, if this conclusion were formed by reason, it would be as perfect at first, and upon one instance, as after ever so long a course of experience.'[3] Combining (1) and (2) we arrive at (3): The inference from past conjunctions, together with the present

[1] Ibid. ch. 4.
[2] Cf. Stove, p. 76.
[3] Hume, *Enquiry concerning Human Understanding*, p. 36 (Selby-Bigge).

instance of flame, is no more conclusive than the inference from no empirical data at all.

Symbolically, (1), (2) and (3) may be expressed as follows:

(1) $P(\text{Hot}(a), \text{Flame}(a)\ \&\ t) = P(\text{Hot}(a), t)$;[1]

(2) $P(\text{Hot}(a), \text{Flame}(a)\ \&\ \text{Hot}(b)\ \&\ \text{Flame}(b)\ \&\ \text{Hot}(c)\ \&\ \text{Flame}(c)\ \&\ t) = P(\text{Hot}(a), \text{Flame}(a)\ \&\ t)$;

(3) $P(\text{Hot}(a), \text{Flame}(a)\ \&\ \text{Hot}(b)\ \&\ \text{Flame}(b)\ \&\ \text{Hot}(c)\ \&\ \text{Flame}(c)\ \&\ t) = P(\text{Hot}(a), t)$.

Putting 'h' for the conclusion and 'e' for the empirical data appearing in the left hand formula of (3), Stove recasts (3) as

(4) $P(h, e\ \&\ t) = P(h, t)$.

This is his formulation of Hume's inductive irrationalism.

Now Stove claims that (4) is inconsistent with a proposition entailed by obvious and fundamental propositions of logical probability. This is the proposition

(J) $P(q\ \&\ r, r\ \&\ t) > P(q\ \&\ r, t)$

(where 'q' and 'r' are contingent).

Indeed, (J) not only follows from intuitively obvious premisses but seems obvious enough itself. In informal terms it states that a conjunction '$(q\ \&\ r)$' is less likely to be false given the truth of one conjunct 'r' than if the truth of neither conjunct is given. For if 'r' is given, one possible source of the falsity of the whole is excluded, namely the truth of '$\sim r$', whereas if neither is given, this source remains open. The suggestion that a conjunction of contingent propositions might be no more likely to be true given the truth of one conjunct than if the latter were not given, would be equivalent to the absurd suggestion that that conjunct itself is no more likely to be true if given than if not given.

With appropriate substitutions, (J) gives rise to

(J') $P((h\ \&\ e)\ \&\ e, e\ \&\ t) > P((h\ \&\ e)\ \&\ e, t)$,

which in turn is equivalent to

(J'') $P(h\ \&\ e, e\ \&\ t) > P(h\ \&\ e, t)$.

Now, according to Stove, (4) above entails

(5) $P(h\ \&\ e, e\ \&\ t) = P(h\ \&\ e, t)$;

which is inconsistent with (J''), hence with (J').

Stove can pass from (4) $P(h, e\ \&\ t) = P(h, t)$ to (5) $P(h\ \&\ e, e\ \&\ t) = P(h\ \&\ e, t)$ only by substituting in (4) '$(h\ \&\ e)$' on 'h'. But is it legitimate

[1] Throughout, 't' is a tautology.

to substitute on '*h*' (the conclusion) a conjunction of which '*e*' (the non-tautological premiss) forms one conjunct?

In taking the substitution to be permissible, Stove reveals that he, like a number of modern writers on induction, understands the expression 'inductive argument' to cover any argument from an empirical premiss '*P*' to an empirical conclusion '*C*' where '*P*' does not entail '*C*'. Thus, if '*C*' consists of '(*P* & *H*)', the argument '*P*, therefore *P* & *H*' is inductive. If arguments of this form are to be classed as inductive, then certainly it follows from fundamental propositions of logical probability that some inductive arguments can be rendered more conclusive by expansion of the premisses, viz. those where the conclusion is a conjunction of independent conjuncts one or more of which is then introduced into the premisses. For the principle embodied in (J) above entails that $P(q \,\&\, r, s \,\&\, t) < P(q \,\&\, r, r \,\&\, s \,\&\, t)$, where '*q*', '*r*' and '*s*' are independent empirical propositions.

But the truth of (J) is not inconsistent with *Hume*'s inductive irrationalism. The inductive ("probable") inferences whose rationality Hume denies are, as his text makes plain, experience-based inferences from one event ("object") to a logically independent event. Nowhere does Hume suggest that he is concerned with inferences to a conjunction one of whose conjuncts already figures in the premisses.[1] In other words, his position is properly represented by (4) above, i.e. '$P(h, e \,\&\, t) = P(h, t)$' where '*h*' is *not* a conjunction one of whose conjuncts is '*e*'. But with this restriction on possible substitutions on '*h*', (4) does not give rise to the absurd proposition (5).

Qua inductive irrationalist Hume says nothing to indicate that he could not accept that a conjunction of contingent independent propositions '(*q* & *r*)' is more likely true given '*r*' than not given '*r*'. This acceptance would be common ground between him and the believers in the rationality of induction to whom his position is supposed to represent a paradox and a threat. What, then, is the difference between Hume and them? The believers are not only confident that '(*q* & *r*)' is more likely true given '*r*' than not given '*r*', on the ground that it cannot then be false on account of not-*r*; they are also confident that for some values of '*q*' and '*r*', '(*q* & *r*)' is more likely true given '*r*' than not given '*r*', *on the ground that* '*q*' (and therefore '(*q* & *r*)') *is more likely to be true given* '*r*' *than not given* '*r*'. It is this latter belief that Hume rejects, and his rejection remains formally unrefuted.[2]

[1] Hume virtually equates inductive reasoning with reasoning from cause to effect (given past experience) and *vice versa*. But the effect (or cause) inferred to does not contain the cause (or effect) as a part of itself. Nor does it contain any of the past experiences on which the inference is founded.

[2] My thanks are due to Alexander Broadie for discussion of this paper.

University of Edinburgh SARAH WATERLOW 1975

THE PROVINCE OF LOGIC

By CHARLES SAYWARD

IN *Ways of Paradox* (New York, 1966) Quine criticizes Strawson's account of the province of logic (pp. 138–9):

> If [Strawson] has chosen too soft and friable a keystone in analyticity, then it is fair to ask what he could have used in its place. Insofar as he uses analyticity in defining the province of logic, my answer is as follows: he could have used, instead, the notion of truth and the notion of logical vocabulary. Given these, the business of formal logic is describable as that of finding statement forms which are *logical*, in the sense of containing no constants beyond the logical vocabulary, and (extensionally) *valid*, in the sense that all statements exemplifying the form in question are true. Statements exemplifying such forms may be called *logically true*.

Robert Hadley ('Quine and Strawson on Logical Theory', ANALYSIS 34.6, pp. 207–8) takes issue with this criticism of Strawson, arguing 'that any application of Quine's definition of logical truth presupposes prior intuitions about what inferences are valid and what truths are necessary' (p. 208). I agree with the first conjunct of this conclusion but contend that it does not adversely affect Quine's criticism of Strawson. I maintain that Hadley's case for the second conjunct is based on a confusion.

First the latter point. Hadley asserts 'Quine's definition is acceptable only if we construe "All statements exemplifying the form in question are true" to mean that all *possible* statements of the form in question are true' (ibid., 208). Hadley argues, persuasively, that the difference between analyticity and necessity is not relevant to the present issue. So the point must be that Quine's account of logic makes implicit use of the notion of analyticity.

There seems to be a pretty obvious reply to this argument. The context makes clear that by 'statement' Quine meant 'sentence having truth conditions'; in particular, Quine did not intend 'statement' to refer to occurrences of sentences. *Prima facie* it makes sense to discriminate actual occurrences of sentences from possible occurrences. But, insofar as 'possible sentence' makes any sense at all, it is coextensive with 'actual sentence'. In this respect, a statement, in Quine's sense, is like a formula of a logical system. A string of symbols of the system, whether actually produced or not, is a formula just in case the string satisfies certain rules. For Quine both a formula and a sentence are set-theoretical entities. (Quine, *Word and Object*, New York and London, 1960, p. 195.) Thus Quine's use of 'statement' involves modality only if set theory does.

The second point concerns Hadley's contention that Quine's definition of logical truth presupposes prior intuitions about what inferences

are valid. His basis for saying this is that 'we could not *prove* that all possible statements exemplifying the form in question are true without presupposing the *validity* of some inference' (ibid., 208). How is this supposed to count against Quine's criticism of Strawson? Hadley's case must go something like this. Quine criticizes Strawson for characterizing the province of logic in terms of analyticity. Quine gives an alternative characterization in terms of truth and logical vocabulary. Hadley maintains that applying Quine's characterization presupposes prior intuitions about what consitutes a valid inference. At this point there are two ways of interpreting Hadley.

One possibility is to interpret Hadley as concluding that Quine's characterization of the province of logic is circular. If so, then Quine would not have succeeded in giving an adequate alternative to Strawson's account. But, interpreted in this way, Hadley's argument is invalid. Surely to prove that a sentence of a language L is a logical truth in the language one would have to supply a valid inference in a metalanguage about L. This does nothing to show that Quine's definition of logical truth or his definition of the province of logic is circular.

A second possibility is to interpret Hadley as claiming Quine's characterization of logic makes implicit use of the notion of analyticity. To get to this conclusion Hadley needs the premiss that analyticity and valid inference are interdefinable notions. But this is really the very point at issue. So, interpreted in this way, Hadley's case is question-begging.

University of Nebraska—Lincoln CHARLES SAYWARD 1975

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, R. G. Swinburne; J. H. Benson, Margaret A. Boden, Andrew Harrison, R. F. Holland, Hidé Ishiguro, J. Kemp, Bernard Mayo, D. H. Mellor, A. R. White, Peter Winch. This committee is responsible for appointing and advising the editor and for the general policy of the paper.

SUBSCRIPTIONS. The subscription to ANALYSIS for institutions is £5.00 (inland), £6.00 (overseas), $15.00 (U.S.A. and Canada); for individuals £4.00 (inland), £4.80 (overseas), $12.00 (U.S.A. and Canada). Each volume comprises four numbers, three of 48 pages and one of 64 pages, appearing within the academic year—in October, January, March and June. Orders should be sent to Basil Blackwell, 108 Cowley Road, Oxford OX4 1JF, or placed with any bookseller.

CONTRIBUTIONS. Articles submitted for publication should be addressed to Dr. C. J. F. Williams, 1 Fossefield Road, Midsomer Norton, Bath, BA3 4AS. Contributors are asked to note the following requirements.

Articles should **normally** not exceed 3,000 words in length. **Occasionally,** however, longer contributions can be accepted;

They must be typewritten in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

It is regretted that owing to increased postal charges it is no longer possible to return typescripts unless the following instructions are followed:

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgment is required, a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope **and international reply coupons** of the requisite value, whether for air or surface mail.

Galley proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts will be retained by the Editor on the assumption that authors have kept their own copies.

PRINTED IN GREAT BRITAIN BY BURGESS & SON (ABINGDON) LTD., ABINGDON, OXFORDSHIRE

Vol. 36 No. 2 (New Series No. 170) January 197

ANALYSIS

Edited by
C. J. F. Williams

CONTENTS

BASIL BLACKWELL · ALFRED STREET · OXFORD

ANALYSIS "PROBLEM" NO. 15

BETWEEN 1952 and 1958 fourteen "problems" were set in ANALYSIS by distinguished philoscphers, and solutions to them published in the journal. It has been decided to revive this competition. The fifteenth problem is set by Professor B. A. O. Williams, of Cambridge, and is as follows:

> There exist just 100 tablets in a certain script, as yet undeciphered. Scholar A considers all the tablets, and works out a decipherment which makes sense of them. Independently, scholar B selects 50 of the tablets, and from them works out a decipherment which he then tries out, successfully, on the other 50. Is one of these scholars behaving more rationally than the other? If so, which, and why? If their decipherments differ, does the difference in procedure give any reason to accept one rather than the other?

Entries (of not more than 600 words) should be sent to the Editor of ANALYSIS by the 30th June, 1976. They should be accompanied by stamped addressed envelopes or international postage coupons, if return of the typescript is desired. No entries should be sent to Professor B. A. O. Williams. Contributors may submit entries either under their own names or under a pseudonym. Contributions can be accepted only from undergraduate or postgraduate students of philosophy.

A report with any winning entries will be published in Volume 37 of ANALYSIS. The ANALYSIS Committee has voted a sum of £40 which will be awarded as a prize if the adjudicator finds a sufficiently deserving contribution.

DUMMETT ON ABSTRACT OBJECTS

By H. W. NOONAN

IN Chapter 14 of his book (*Frege: Philosophy of Language*, Duckworth, London, 1973) Michael Dummett offers an account of the distinction between concrete and abstract objects. I think that Dummett's account contains a flaw it will be illuminating to expose.

After some preliminary discussion of the difference between a Fregean and an Aristotelian outlook and Quine's criterion of ontological commitment Dummett looks first at the suggestion that the distinction between concrete and abstract objects is just the distinction between those objects which are, and those which are not, accessible to the senses. He rejects this suggestion, partly because of its vagueness, but goes on to

consider a distinction with some affinity to it, but of more philosophical significance—the distinction between those objects which are, and those which are not, possible objects of ostension.

A pointing gesture by itself, of course, cannot pick out an object—one does not identify something simply by pointing and saying 'This'—one has to be able to answer the question 'This *what*?'. The explanation of this necessity, according to Dummett, is that when one picks out an object by means of a demonstrative either the context must supply an appropriate criterion of identity, or one must expressly be given by means of a general term; otherwise no definite object is indicated at all. But in some cases the answer to the question 'This what?' will be insufficient to identify the object being talked about. If I point at a red expanse and say 'This colour' this makes quite clear what I am pointing to. But if I point and say 'This shape' this does not yet make clear what I am pointing to. I have to identify the shape I am pointing to by saying of what object or region it is the shape. If I am to be in a position to identify shapes and so to use proper names of shapes I therefore need to understand some functional expression with the sense of 'the shape of'.

Dummett consequently puts forward for examination the suggestion that objects of a certain kind are abstract if and only if there is some functional expression such that it is essential to the understanding of any name of an object of that kind that the referent of the name be recognised as lying within the range of that functional expression. On this criterion shapes and numbers come out as abstract objects, while colours come out as concrete objects. Dummett notes, however, that this is only so because the criterion is construed strictly, as exclusively concerned with logical necessity. One's ability to name colours, like one's ability to name shapes, is, as a matter of fact, and probably as a matter of psychological necessity, acquired after one's ability to name physical objects which possess colours and shapes, and along with one's understanding of a functional expression with the sense of 'the colour of'. But there is no logical or epistemological absurdity in supposing the understanding of names of colours to be acquired before the understanding of names of physical objects and so in the absence of any understanding of any functional expression with the sense of 'the colour of'. This is because, as we have seen, colours, unlike shapes, are possible objects of ostension.

However, Dummett decides that, while this suggestion does give a sufficient condition of something's being an abstract object, it does not give a necessary condition. The reason is that he wishes to regard conventions in Bridge and Chess openings, and indeed Bridge and Chess themselves, as abstract objects, but none of these falls within the range of any functional expression of the kind under consideration. (The reason for this, in turn, as we shall see, is that all these can be picked out by ostension, but Dummett does not recognize this.)

Now I shall present my criticism of Dummett's position. It is this: not only does the suggestion under consideration not provide a necessary condition of abstractness, as he notes; it does not, contrary to his assertion, even provide a sufficient condition.

Consequently, the line between those objects which are possible objects of ostension and those which are not cuts right across the concrete/abstract line.

It is really rather easy to see this. Colours could be named by a people who had no understanding of any functional expression with the sense of 'the colour of' and, moreover, had no means of naming physical objects such as have colours. Shapes could not be named by such a people. Whether or not names of colours are temporally prior to names of shapes in the acquisition of language, they are certainly *logically* prior. And the reason, as we have seen, is simply that colours can be identified just by pointing and saying 'This colour', whereas shapes cannot be so identified. But what, in turn, is the reason for this? It is simply that if an object possesses some one colour then every part of that object which possesses a colour (not counting parts too small to possess colours) will possess the *same* colour as the object of which it is a part. So the direction of the pointing finger will pick out just one colour. However, an object or region which has just one shape (as every object or region has, of course) will also have proper parts which are of some shape or other, but not of the same shape as it is. So if I point to an object and say 'This shape' nothing is yet clear. I will be pointing to ever so many objects of different shapes. I must, then, further identify the shape I mean by saying of what object it is the shape.

But consider: suppose I point and say 'That gold' or 'That water'—will I have identified a parcel of gold or water? Obviously not: in order to do so I will have to specify which region the parcel of gold or water I am intending to point to occupies. And the reason why this further specification is necessary is the same as in the case of the shape: any region which is occupied by a parcel of gold or water will contain smaller regions, also occupied by parcels of gold or water, but by different parcels. (At least, this is so at the macroscopic level, and it is not excluded by the *senses* of the expressions '(parcel of) gold', '(parcel of) water' that it should remain true at the microscopic level also.) The only difference is that while a proper part of an object of a certain shape can be the same shape as it, a proper part of a region occupied by a certain parcel of gold or water cannot be occupied by just the same parcel of gold or water as it is.

One cannot, then, identify either shapes or parcels of gold or water in the same way as one can identify colours because of a feature shared by the relations *being the same gold, being the same water* and *having the same shape,* but lacked by the relation *having the same colour.* It is because of this

feature that 'gold', 'water' and 'shape', unlike 'colour', are not count nouns. (Of course, these last remarks ignore the scientific discovery that there are, in fact, smallest parcels of gold or water—atoms and molecules —so that it is, in principle, possible to count parcels of gold or water. But, because of the compoundability of parts that are gold or water into a whole that is also gold or water, it is still impossible—even in principle —to pick out a parcel of gold or water by ostension in the way one can a man or a colour. This shows that the uncountability of things of kind K is not a necessary condition of K's not being possible objects of ostension but it is certainly a sufficient condition, and that is all I need for the validity of my criticism of Dummett.)

Similarly, the reason why games or chess openings are possible objects of ostension is simply that one can count how many games are going on in a particular region at a particular time or how many chess openings are being played there then (if ten people are playing chess, twenty bridge and two snakes-and-ladders, that makes three games; if the Ruy Lopez is being played on two boards, the Sicilian Defence on two and the Giuoco Piano on one, that makes three chess openings), while, of course, games or chess openings are not compoundable into other games or chess openings.

But now, if all this is correct, what are we to say of Dummett's worries about the reality of such objects as shapes, or his diagnosis of the origin of nominalism? Before considering these questions let us first look at his treatment of the idea that an abstract object cannot be involved in causal interactions with other objects. He attempts to justify this idea by pointing to the 'non-explanatory character of any statement that can be made about the abstract object in itself'. He continues:

> If, for the sake of argument, we accept the popular belief that a red rag infuriates a bull, are we to say that the colour causes the bull to charge? There seems no especial reason why not: but this is because we are not regarding the colour as an abstract object and, therefore, allowing it to have a spatial position, contingently. We can explain the bull's rage by the fact that the colour was there, where he could see it. Contrast the theory that the taste of a substance is determined by the shape of its molecules; could we say that a certain shape causes a bitter taste? In so far as we regard a shape as a genuine abstract object, and, therefore, as not having, in itself, a spatial position, but merely as enjoying the property of being the shape of this or that object or configuration, we are reluctant to say this: the taste resulted, not from the presence of the shape, but from the presence of a molecule of that shape. This reflects the fact that shapes are *of* objects in a way that colours are not: it is not merely that we do not choose to say that a shape is in a particular place, whereas we more readily say this of a colour, but that it would make no sense to say that a shape was in a place without giving some indication of what it was the shape of. A point has, indeed, spatial location, but in the sense in which a point cannot move, not contingently so. To give a cause of some occurrence we must cite some contingent facts. (This truistic principle can be

> accepted without any deep analysis of 'contingent' since all we need is its intuitive sense of 'something that might have been otherwise':..., it is platitudinous that statements of causality yield counterfactual conditionals.) No contingent fact about an abstract object can be cited that cannot more naturally be construed as a fact about concrete ones, for instance, the concrete object which the abstract one is 'of': and hence we do not regard abstract objects as being themselves causally efficacious or the subjects of causal effects. (Op. cit., p. 493.)

Obviously, something has gone wrong here. It is true that we do not think of shapes as involved in causal interactions with other objects. But the explanation of this fact cannot be the one Dummett gives—namely, the way in which a shape is 'of' a concrete object. For the feature of the notion of a shape by means of which he explains this idea is the one we have been discussing, which is equally a feature of the notion of a parcel of matter (water, gold), but, of course, we do think of parcels of matter as involved in causal interactions. It is something which *distinguishes* parcels of matter and shapes that needs to be pointed to in explaining why shapes cannot be thought of as involved in causal interactions. Moreover, even if Dummett's account had sufficed to explain why shapes cannot be thought of as affecting and being affected by other objects, a quite different explanation would have been needed of why such an abstract object as Chess, which is a possible object of ostension, cannot be thought of in this way. But surely, if the distinction between abstract and concrete objects is of any significance, then the explanation of this inconceivability, for any abstract object, ought to be the same as its explanation for any other.

Let us now consider Dummett's worries about the reality of objects which cannot be objects of ostension. From viewpoints as distinct as those of Aristotle and Russell these worries have to be regarded as justified. Aristotle and Plato both used the expression '*tode ti*'—'this something'—and it played an important part in their philosophies, especially in Aristotle's. For Aristotle a substance was a 'this something', *to be* in the primary way was to be a 'this something'. When he came to think of the Form rather than the particular as substance, Aristotle also transferred his application of the expression '*tode ti*' to Forms. Now, of course, something which is not a possible object of ostension fails, in the most obvious way, to be a 'this something'—such an object can be identified only by purely linguistic means, or by identifying some other object by ostension and identifying it as *the* object of its kind which is 'of' the ostended object. Again, as is well known, for Russell the fundamental constituents of reality were those things that could be named by logically proper names, and, as is equally well known, Russell eventually came to think that the only logically proper name was 'this'. From this viewpoint too, then, things which cannot be picked out by ostension have to be regarded as in some way ontologically inferior. So

I do not wish to say that these worries of Dummett's are unjustified, but only that they are in no way worries especially about abstract objects, and so are not relevant to nominalism.

Similarly, when Dummett goes on to characterise nominalism as prompted by the feeling that one cannot be *shown* an abstract object, the way he interprets this statement makes it true also that one cannot be *shown* a parcel of matter. Nevertheless, the statement that one cannot be shown an abstract object is, in itself, one that we should intuitively wish to accept, and does doubtless lie behind the nominalist position. The problem is to interpret it in some other way than Dummett's, since his way, as we have seen, is inadequate. Obviously our intuition here is connected with our intuition that abstract objects cannot be involved in causal interactions with other objects—to see an object, after all, *is* to interact with it causally. But I shall leave this matter here.

This discussion has mostly been negative, but I think there is something positive to be learned from it. We have seen that the distinction between colours and shapes which Dummett is concerned with is based upon, and explained by, a distinction between the relations *having the same colour* and *having the same shape*, namely, that the first but not the second of these relations is one in which two things, one a proper part of the other, must necessarily stand to each other. The fundamental distinction, then, is not between kinds of object, but between kinds of relation. Now, if we agree with Dummett that the introduction of a name into a language requires its association with a criterion of identity, and that the explanation of the criterion of identity associated with a name constitutes the most important part of the explanation of its sense, then we must surely also insist that it is not possible for there to be two names associated with the same criterion of identity, one which names an abstract object and the other a concrete one. Well,—and this is my suggestion—a way to ensure this result would be to explain the abstract/concrete distinction between objects as derivative from a distinction between relations which can serve as criteria of identity for names of objects. The provision of such an account would doubtless be a task of great difficulty, but it seems to me—and I think ought also to seem to Dummett—that this is the only correct way to approach the abstract/concrete distinction.

Trinity Hall, Cambridge

MISCONCEPTIONS ABOUT MORAL NOTIONS

By ROGER A. SHINER *and* JEROME E. BICKENBACH

IN 'Moral Notions and Moral Misconceptions' (ANALYSIS 35.3, 65–78) Keith Graham states his lack of sympathy with Julius Kovesi's analysis of moral notions in his book of that name.[1] Although, like Graham *vis à vis* Kovesi, we find the former's arguments 'lively, penetrating and original', we are equally unsympathetic to *them*. We believe that Graham has not understood what a Kovesian view involves, and that such a view is not defective in the way he suggests. We use the expression 'Kovesian view' deliberately. We do not want to get bogged down in fundamentalist hermeneutics of *Moral Notions* as a canonical text; we want rather to discuss the issues themselves. We do believe that the text importantly illuminates these issues, and will provide some chapters and verses which we hope will satisfy the fundamentalists. However, our aim, as Graham's, is broader.

I

In the first part of his paper, Graham criticizes Kovesi's doctrine of material and formal elements, and expresses doubts about the value of these notions for our understanding of morality. We shall try to show that these criticisms are misplaced.

Kovesi introduces the terms (3, ff.) in part to bring out what he thinks is a very misleadingly expressed insight behind the emphasis on "Hume's Law" in recent moral philosophy. In the later positive part of his paper Graham produces his own account of what lies behind "Hume's Law". To begin with he tries to show that Kovesi has not succeeded in bringing this out any more than has Hare or any other recent orthodox (as it were) scholar. Kovesi wants to agree with the orthodoxy that indeed one cannot move from 'Is' to 'Ought' as the orthodoxy understands that. However, he denies that the explanation has anything to do with a special problem about the nature of moral reasoning. Rather, he represents it as a special case of the general difficulty of moving from the material elements of a notion (*any* notion) to the formal element by a principle of entailment. This move Kovesi takes to be impossible for any notion, not just peculiarly moral notions (cf. e.g. 11–12). But he does not think that all kinds of evil scepticism follow from this, for he denies this inability has anything to do with the availability of the requisite kind of knowledge. The inability arises because the idea of making such a move is incoherent. We do not have available an antecedently existing and open stock of material elements on which we draw when we want to construct a notion.

[1] *Moral Notions* (Routledge and Kegan Paul, London, 1967).

Rather, that such-and-such are the material elements of notion N is something we analyse out of notion N when it has become established. The Aristotelian overtones of Kovesi's terminology are, we imagine, quite deliberate. For instance, Aristotle thinks of the body as the matter and the soul as the form of a living creature.[1] But he denies that a body is a body-of-a-living-creature until it has been informed by a soul. There is no Great Artificer putting together bodies and souls to make living creatures; we analyse the elements out of an actual living creature.

Graham tries to pooh-pooh all of this by pointing out that (i) it is necessary for a thing to be a table that it has some shape, and therefore we have a necessary condition for being a table; also (ii) that in the case of a description of some particular table, that description is surely sufficient in that case for something being a table, even though not necessary. Either of these claims, if true, Graham implies, will shed doubt on Kovesi's line on non-entailment.

Graham's first claim is undoubtedly true, but it misses the point Kovesi is making. Kovesi wants to argue that the problem of connecting what 'is given to the senses' to 'what we judge a thing or act to be' presents itself in the same way for both descriptive and evaluative terms (10–12). He notes that, in the case of material objects, the 'gap' is viewed as something of a perennial problem in the theory of knowledge; but we should not view the problem as one of finding an entailment between what is given to the senses and what we judge the thing to be. Likewise, Kovesi argues, given the parallel, there is no need for moral philosophers to be condemned to search for an entailment between "descriptive" terms and evaluative terms. Kovesi is here pointing to the confusion involved in aligning description with 'what is given to the senses' and evaluation with 'what we judge a thing or act to be'. Now, clearly, Kovesi sees this parallel because he is giving the phrase 'what is given to the senses' the traditional reading, namely to mean 'appearances', 'sense data' and the like. On this reading, it is *not* the case that it is necessary for a thing to *be a table* that it *appear to have some shape*. Kovesi's point is that *this* lack of connection poses a problem for the application of descriptive terms. Thus the necessary condition Graham points out is not the one Kovesi is trying to deny.

With respect to the second claim about sufficient conditions, Graham correctly diagnoses the Kovesian reply—that it is always possible for there to be a concept like *table** which is similar in material elements but different in formal elements from ours, and that moreover the formal element is what groups the material elements together in the first place. However, he does not appreciate what is being argued here. In order for it to be the case that an enumeration of elements entails 'This is a table'

[1] Cf. *de Anima* II 1–3, esp. 412 a 28, ff.; also G. E. M. Anscombe and P. T. Geach, *Three Philosophers* (Blackwell, Oxford, 1963), 54–58.

in a way problematic for Kovesi, we would have to be able to move from that list *raw straight* to the desired conclusion, and not dog-leg via some assumed social convention. The point of the *table** story is to deny this. Such a counter-example need not appear trivial, if a full enough story is told. Suppose in fact that the table* people use tables* by turning them upside down (as we would say, though to them "upside down" is right way up) and have races downhill over snow on them. Or suppose that they use them only for certain ceremonial purposes—say, school graduands are carried on them to receive their degrees. Certainly, *we* can say, 'Their tables* could be used as tables', and doubtless *they* could say the reverse; but that is beside the point. Also, there are here as well other possible cases—suppose they turned tables upside-down, cut notches in the legs, and used them for winding wool. A table with notches in the legs would still serve the purpose of a table. Yet the inclination might be here to deny more strongly that tables* were tables, because of the notches. The thing to see is that the cases of tables* we have described are the limiting cases; but they still prove Kovesi's point.

By contrast, as soon as one does dog-leg *via* the assumed convention, then indeed entailments can take place. That is, 'Given our needs and social conventions, then "This consists of a flat mahogany surface resting on three iron legs" (or whatever) entails "This is a table" ' is true. The point is, such a "given" is assumed in *every* case of entailments of this kind. Therefore, that it is required in the moral case does not *eo ipso* mark anything peculiar or problematic about morality.

Graham's move now is to complain about Kovesi's idea that formal elements collect together empirical features into notions. He says that Kovesi does not make it clear what is to count as an empirical similarity/difference. To a degree this complaint is justified, for Kovesi indeed doesn't give an account of this term. He mentions examples, that two things the same shade of colour are empirically similar, and that knocking over a cup and tripping over a sandcastle are empirically dissimilar. But this need not be a vicious deficiency, unless it turns out that attempts to articulate the point will be question-begging. This in fact Graham claims to be the case. He is prepared to countenance empirical similarities which are 'detectable at a higher level of abstraction' (67); given this concept of *empirical similarity*, then two acts of murder or of inadvertency *will* be empirically similar, as will two patches of heliotrope. Thus, maintains Graham, we can analyse the concept of murder in terms of empirical similarities if we recognize the existence of highly abstract empirical similarities.

This seems to us to be a merely 'verbal manoeuvre'. One feels sympathy towards Graham's claim that acts of murder are empirically similar at a higher level of abstraction (but then of course *any* two things are similar at *some* level of abstraction). Is this not precisely because one sees

that they are all instances of the concept *murder*, that they all amount to murder, that one who knew what it was to be a case of murder would recognize these to be cases of murder, and so forth. That is precisely what the Kovesian view is here claiming. What Graham needs is this—that simply from being faced with these cases alone, with no knowledge of human needs and social conventions, a Martian (or whoever) would be able to abstract from these cases that they were all cases of a certain concept; that he would be able to do this when faced merely with observational (in a narrow sense) evidence. That is surely monstrously implausible. What is right here, and rightly emphasized by Graham, is that *red* (e.g.) and *murder* (e.g.) are on a kind of continuum. In order to be able to recognize instances of *red* as empirically similar one does need to be aware of some convention—but merely the conventions of naming colours. Progressively more complex conventional understanding is needed to grasp *murder*. But when Graham claims that cases of *murder* are empirically similar, he makes a double mistake. He makes the mistake of thinking that recognition of *red*, e.g., operates independently of a Kovesian convention, and of thinking that the difference of degree between *red* and *murder* is not a relevant difference.

Our aim in this section has been to show that Graham's objections to Kovesi are not based on non-theory-laden descriptions of the cases he brings up. Rather, his objections are still couched in the terminology and concepts of the position Kovesi is attacking. We grant that this does not establish the truth of the Kovesian view. But it does clarify the commitments of that view, and show them to be less problematic than Graham maintains.

II

In section II of his paper, Graham begins with two important arguments, the purpose of which is to show that Kovesi's notion of 'formal element' is confused and cannot do the work which (indeed) is required of it. We will deal with the second one first; the first one is picked up again later in the article, and we will deal with it at its second occurrence.

Graham's second important argument also picks up something he has said earlier. He remarked at the end of section I that

> It might be that there is no further activity for the sake of which the given discrimination among geometrical figures is made: the point in discriminating and classifying tak-shaped figures may simply be to discriminate and classify them (67).

The implication he wishes to draw is a confusion in Kovesi's concept of *formal element*. The alleged confusion is then explicated in his section II by means of the distinction between P-type and C-type answers to the questions 'Why do we call certain things cases of x?'; these answers are in terms of purposes and common properties respectively. Kovesi claims

that both supposed misuses of the formal element of a notion and supposed misuses of the material element of a notion are decided ultimately by appeal to the formal element (59–60). This is part of his strategy for arguing for the fundamental position of the formal element. Graham's counter-strategy is to argue that the appeal to the formal element in the case of a supposed misuse of the material element amounts to giving a *C-type* answer, in the above sense. He claims that, if this can be traced back to a formal element at all, it can be so only in a quite different sense of 'formal element', namely, one in which the purpose is to give the empirical common properties according to which a term is ascribed. The conflation of these two types of formal element by Kovesi is, Graham argues, particularly damaging in the case of moral notions, for moral disagreement often arises despite agreement on the appropriate C-type answer.

Again, we agree that this is a correct account of a distinction that a Kovesian view can absorb, and indeed should absorb. Again too, however, it seems to us that in thinking recognition of this distinction is inconsistent with a Kovesian view Graham has simply misunderstood the latter. He is still looking at the Kovesian view through the spectacles which Kovesi is trying to remove from the noses of moral philosophers (cf. Wisdom 'Metaphysics and Verification' *Mind* 1938, p. 498). In various places Kovesi recognizes that there is such a thing as the descriptive point of view.[1] Certainly, what we are up to in applying terms from this point of view is not what we are up to when applying terms from the evaluative, or moral, point of view. In the case of the descriptive point of view, the material elements will be very closely connected with the formal element, in the way Graham's C-type answers imply. The mistake, however, is to think that P-type answers and/or the moral point of view are in any sense *superimposed* on C-type answers or on the descriptive point of view. Kovesi, in a pair of vivid and pertinent culinary metaphors, refers to the view he is attacking as the 'icing' view of moral notions (25)—there is an antecedently existing world of description and moral evaluation is the icing on that world. This view however confuses two questions—'The first question is like asking how we make scrambled eggs, fried eggs or omelettes out of eggs, and the second is like asking how to make scrambled eggs out of fried eggs. But just because we never make scrambled eggs out of fried eggs it does not follow that we do not make both out of eggs' (63–64). Raw eggs are the "material elements" for fried eggs and scrambled eggs (in the analogy, legal (e.g.) and moral notions). We can't indeed make fried eggs out of anything except raw eggs, nor can we make scrambled eggs out of anything else either. But raw eggs are not *metaphysically raw*; 'raw egg' is a notion with *its own* material and formal elements, even though *its* formal

[1] Cf. e.g. 63, 86, ff., 161 and 146–148 especially.

element (in this context) amounts to little more than what enables a cook to pick this ingredient out of his store cupboard and not some other one. The point of view here is descriptive or classificatory; however, the descriptive point of view exists *alongside* the moral point of view, and *not* as something prior to it or more fundamental than it.

Graham does think C-type formal elements are prior—'the point is to condemn *a certain kind of action* which can be discriminated from other kinds of action only by means of a C-type criterion' (71; his italics). But *this* procedure of discrimination is, for Kovesi, carried out by means of 'recognitors'—'the defining characteristics of the material elements of a thing or act or situation or any phenomenon' (41; cf. also 46, 54–55). The priority which exists here is a temporal one only. Certainly we have to recognize what we have in front of us before we can do anything else. But the question 'What makes that a case of kind K?' is not the same as 'How did you recognize that to be a case of kind K?'. The priority and centrality of some feature for the latter question does not entail its priority and centrality for the former. The former issue is decided on quite different grounds. Moreover, that 'someone may . . . agree with us on what is to count as another instance of the same thing but disagree with us with regard to condemning it' (Graham 71) is a dogma of the 'icing' view Kovesi is attacking. The key word omitted is 'may *legitimately* disagree'. If *this* is so, then for Kovesi the moral notion concerned must therefore be incomplete, and we have a moral disagreement to be resolved by the appropriate means. Of course, if the claim being made here is that anyone may always legitimately disagree over the application of any moral term to anything, then, if that claim is true, the Kovesian view is false. But this can be only a conclusion of the dispute between Graham and Kovesi, not a permissible move in it.

An illuminating analogy here is with a feature of Austin's view in *How to Do Things with Words*.[1] Austin's insistence there that the fundamental unit of language is *the total speech-act* (147) has much in common with the Kovesian view. Austin takes his view to 'play Old Harry with the fact/value fetish' (150), and indeed it does. In the same way as Kovesi, Austin is combating the view that description exists antecedently to evaluation, and the central problem of moral philosphy is to understand the move from one to the other. Instead, Austin argues, describing, stating, evaluating, condemning, etc., are all different kinds of speech-act, differing in illocutionary force. Describing is a speech-act which co-exists alongside evaluating; each has its own set of conventions. Inside of each is the locutionary act, the act of using words with a sense and reference as words with a sense and reference; there are analogies here with Kovesi's material elements. But no speaker can ever perform merely a locutionary act; he must always perform such an act in

[1] Oxford University Press, Oxford, 1962.

performing an illocutionary act (98–99). Likewise, for Kovesi, as we have explained, the material elements do not stand alone, but only as a component recoverable by analysis of a notion.

The Kovesian view is therefore perfectly well able to survive the proof that describing exists as an independent activity with its own conventions. What is at stake here is the relation between describing and evaluating. The Kovesian view does not deny the existence of description: it gives a different account of how description is related to evaluation. The two kinds of formal element represented by Graham's P-type and C-type answers are not merely homonymously called 'formal element'. They are both the same type of formal element; they are relevantly different tokens of that type. For Graham to think that he has caught Kovesi in equivocation and confusion here is merely a demonstration that he has not understood how deep Kovesi is tunnelling.

III

We may now move to Graham's first argument against Kovesi's notion of 'formal element'. This argument, which Graham takes to be a knock-down argument, is that Kovesi's view has the following absurd consequence, that one who wonders whether murder is really wrong, even though he agrees with us on what murder is, is guilty of a conceptual error. This, says Graham, trivializes a genuine moral disagreement. One is reminded here of the character George Moore, a Professor of Moral Philosophy, in Stoppard's *Jumpers*; he remarks of the Professor of Logic, McFee:

> In simple terms he believes that on the whole people should tell the truth all right, and keep their promises, and so on—but on the sole grounds that if everybody went around telling lies and breaking their word as a matter of course, normal life would be impossible. Of course, he is defining normality in terms of the truth being told and promises being kept, etc., so the definition is circular and not worth very much, but the point is it allows him to conclude that telling lies is not *sinful* but simply anti-social (Stoppard's italics).

Graham thinks Kovesi doesn't allow us to think one who regards murder as right is *sinful* but only unconventional or conceptually confused. He glosses our situation when faced with such a man as having first to teach him our language before we can bring out a substantive moral disagreement. Of course, if that were a correct analogy, it would be grotesque. But it isn't correct. A disagreement over the correct application of the concept *murder* just *is* a substantial moral disagreement and *vice versa*. Likewise, a misapplication of the concept *morally justified* just *is* a substantial moral error and *vice versa*. Those who want to do moral philosophy seriously (with *angst* and so forth) are inclined to think that conceptual analysis must be vacant and sterile, and that conceptual

disagreements then must be too, for they are disagreements which conceptual analysis solves. This is however an equally grotesque interpretation of conceptual analysis. What else do the great profound works of philosophy, history and literature give us except a clearer view of the conceptual moral landscape, of what is possible and how things are in the realm of morality? If someone disagrees with us in his moral view, then he understands moral concepts and notions differently from the way we do and acts differently. Two quite illegitimate appeals are covertly being made here. One is of the caricature archetypal moral philosopher trying to convert the Godfather by reading to him *Principia Ethica*. But of course it does not follow from moral error's being conceptual error that this is all we can do. The other appeal is to some form of Subjectivism. For if moral error is conceptual error, then we are not free to choose our moral values as we wish. But for Graham or anyone to have this in mind when they characterize the Kovesian view as 'grotesque' is, of course, hopelessly question-begging.

On page 72 Graham links his criticism of Kovesi on this issue with the latter's category of *complete* moral notions. Part of Graham's strategy is to deny that murder is complete. In passing, let us note that the onus of proof here has to be on Graham to construct a case of an act A which (i) is such that all the material elements of A are material elements of *murder*; (ii) is, moreover, an instance of murder; but (iii) is *not* a morally wrong act (i.e., is either a morally right act or a morally neutral act). If there were such an act A, then Graham would have all that he needs to refute Kovesi's claim that murder is a complete moral notion, for it surely would be the case that here the correct judgement to be made is that act A is a case of murder which is not morally wrong. However, Graham does not supply such a case, and until he does, Kovesi's claim stands.

All the same, little turns on whether *murder* really is a complete notion, or indeed whether any other moral notion is. Moral judgement would still have to be exercised by the responsible moral agent in applying moral notions. A more interesting and more substantial objection is Graham's claim that there ought *not* to be complete moral terms. Someone who claims that murder is sometimes not morally wrong cannot be accused of committing any linguistic or conceptual error, Graham claims; rather than refusing to understand such a man it would be better to argue with him (72). It is important to note that Graham is directing his arguments against the notion of *moral* completeness, rather than completeness itself; for surely Graham would agree that we *should* refuse to understand the man who claims that sometimes even numbers are not divisible by 2. What, then, is there about morality which makes unacceptable the notion of morally complete terms?

Graham's answers seem to be that morality deals with human life and

actions and these things are so complex that it is inadvisable to suppose that particular actions can be adequately and completely described as clear instances of general moral terms. To be sure, Graham argues, murder is *prima facie* wrong: there is some general characteristic of the action-type 'murder' which makes it wrong. This sort of wrongness Graham calls wrongness_G, it is a sort of moral wrongness which we attribute to action-types and which is imparted to particular, concrete action-tokens of these types. But, to say that a particular action was murder, and hence wrong_G, is not necessarily to say that it was a morally wrong thing to do, or in Graham's terminology that it was wrong_D. The wrongness_G of a particular action does not entail the wrongness_D of that action because human actions can be correctly described in an indefinite number of different ways. Thus, an action which is wrong_G may very well be right_D. Because of the complexity of human actions, and the indefinite number of correct descriptions of each action, the gap between correctly describing an action as an action which is *prima facie* wrong (wrong_G) and judging that action to be a morally wrong thing to do (wrong_D) is not a gap bridged conceptually. The person who sincerely claims that murder is sometimes not a wrong thing to do is not guilty of conceptual confusion. He is merely pointing out that it is possible that there be actions which can be correctly described as acts of murder and which can also be correctly described in other ways which would force the judgement that they, despite their *prima facie* wrongness, are not wrong things to do.

The plausibility of Graham's account depends on the sort of case one is considering. In those cases where incomplete moral terms could be correctly applied, what Graham has to say differs only terminologically from Kovesi's account. Cases of, say, lying are *prima facie* wrong, but 'lying' nonetheless describes an action which may or may not be wrong—the term does not fully specify all that we need to know to make a determination of moral rightness or wrongness. If in a particular case, an alternative correct description of an action also correctly described as a case of lying is that of 'saving a life', then the *prima facie* wrongness of lying is overriden by the *prima facie* rightness of saving a life, and the action itself is, *ceteris paribus*, judged to be the morally right thing to do in that case. Thus for incomplete moral terms Graham is not so much countering what Kovesi says as restating it. The difference between the two accounts comes out fully in the case of complete moral terms, as one might expect.

The difference is this. Kovesi argues that one has no further to go once one has determined that the correct description of an action is that it is murder. Graham argues that one *may* have to go further in order to determine whether or not it was the morally wrong or right thing to do in this particular case. Graham thinks this is so, once again, because, according to him, human actions 'can be truly described in an indefinite number

of different ways' (75). In the case of 'lying' this claim underscores the fact that sometimes more must be said about a particular instance of lying to justify the judgement that it was the wrong thing to do in those circumstances. Kovesi's claim that some moral notions like *murder* are complete may be viewed as the claim that if an act can be correctly described as murder then any other description would be, *morally speaking, a misdescription.* Thus, if 'Jones murders Brown' is a correct description of what Jones does, then it would be a *moral* misdescription to suggest that another, correct description of what Jones does is 'Jones (merely?) eliminates a business rival, Brown', even if this latter is a correct description of what Jones does from the descriptive or classificatory point of view. Likewise, if one of the consequences of Jones's murdering Brown is that Jones's impoverished family is provided with unlimited funds from Brown's life insurance policy, it would be a moral misdescription to say that Jones was providing for the financial security of his family. But, on the other hand, if a correct description of Jones's action is 'Jones protects himself from Brown's murderous assaults by killing Brown', then this would surely not be a moral misdescription of the action; but 'Jones murders Brown' then would be a moral misdescription. In other words, even if it is true that there are always an indefinite number of correct descriptions of human actions, still this does not guarantee that each of these descriptions has equal moral weight, i.e., that each of them is morally speaking a fair description of the act.

A Kovesian view is therefore, even if there are or it enjoins complete moral notions, able to absorb the valuable normative recommendation that one argue with people over moral questions. This view does not regard 'moral' and 'descriptive' as incompatibles, nor 'debatable description' as a contradiction in terms. The most serious moral argument can take place over whether 'This is x' is a correct description from the moral point of view, even when x is a complete moral notion.

IV

We shall close by discussing Graham's claim to have brought to light an important kind of moral freedom which he thinks has been obscured by recent writing, and by Kovesi as much as anybody. Again, we do not contest Graham's claim that there is an important point about moral judgement to be brought out here. Neither do we contest his claim that Kovesi has not said as much about this as he might. We do not think the point has received little attention, however, nor do we feel Graham has said what needs to be said.

Graham distinguishes three moves involved in reaching a moral decision—from "factual" considerations (his scare-quotes) to secondary moral considerations, from secondary moral considerations to conclusions involving primary moral concepts in their G-subscript use, and

from here to a conclusion involving a moral concept in its D-subscript use. Graham is prepared to accept a Kovesian account of the first move; he just doesn't think a Kovesian account is much of an improvement over a "traditional" (e.g. Harean) account. We will not remark further on that. The third move is the one Graham thinks has received little attention, and one on which he thinks Kovesi as much as anyone fails to shed light.

On the historical point, Graham seems to us plainly mistaken. His G-subscript/D-subscript distinction seems to us exactly that drawn by Ross between *prima facie* duties and duties *sans phrase*[1] and by Melden between obligation-meeting actions and obligatory actions.[2] Moreover, the distinction seems made for very similar reasons; Ross, for example, writes:

> No act is ever, in virtue of falling under some general description, necessarily actually right, its rightness depends on its whole nature and not on any element in it (33).

He contrasts here moral notions with geometrical notions, of which the above cannot be said. One can even plausibly argue that the point goes back to Aristotle's differentiation of moral generalizations as 'for the most part' rather than 'universal', and of his account of the practical as opposed to the theoretical syllogism. This all seems to us a valid anti-Absolutism in moral philosophy; Graham is here the next in a long line, and not an originator.

Simply characterizing this view as anti-Absolutist does not however deal with the issue of moral freedom. Graham seems to us distressingly unclear on just why he thinks that the kind of freedom he emphasized is 'rather different from the "moral autonomy" talked about *ad nauseam* in recent years' (65). It seems to us he is not so very different from the rest, and that the central distinction here is brought out no better by Graham than by anyone else. The recent talk about freedom and reason (*sic*) has revolved around "Hume's Law" and the impossibility of deriving moral conclusions from factual premisses. The crucial distinction here is between (i) 'freedom' as a term in moral psychology, as something denied to exist by Determinists, and (ii) 'freedom' as a term in the theory of moral reasoning, functioning as a label for the "Hume's Law" feature. That these are fundamentally two quite separate notions is clear, for the standard discussions of freedom and determinism say nothing about Hume's Law. Hare is responsible for the idea that they are connected, for he has taken the view that the first kind of freedom will be secured only if the second kind exists. This is straightforwardly confused, because we are just as free or not free, as the case may be, in the psychological

[1] Cf. W. D. Ross, *The Right and the Good* (Oxford University Press, Oxford, 1930) 19, ff.
[2] A. I. Melden, *Rights and Right Conduct* (Blackwell, Oxford, 1959), 9, ff.

sense, to conclude 'This is a triangle' (to say the words, decide that this is what we are going to judge) given 'This is a plane figure with three sides and three angles' as we are to pass any moral judgement given a "factual" description.

Alternatively, if one considers the talk of authenticity, freedom, existential uniqueness, etc. which, presumably (unless '*ad nauseam*' was a Freudian slip) Graham has in mind, a good deal of this *is* derived from the non-deducibility of 'ought' from 'is'—cf. again Sartre's notorious case of the young Frenchman. We do not see anything in what Graham says that leads us to think his concept of freedom derives from anything else but this feature.

All participants to this debate assume that psychological freedom is a *sine qua non* for making sense of morality; we are no different. The point we wish to emphasize, which cuts against Graham's complaint about Kovesi's complete moral notions, is that the issue of the objectivity or subjectivity of moral notions is quite independent of the issue of the existence or non-existence of this freedom. The conflation of these two issues in recent moral philosophy has hindered and cramped attempts to deal with the objectivity issue, and to us Graham is no less guilty than anyone else. Certainly, insofar as there are complete moral notions, then whether some given particular act is right_G or wrong_D *is decided.* We wouldn't be free to change that, or rather the freedom we would have to change that is no more and no less than the kind of freedom we have to change any of our ways of thinking (what kind of freedom *that* is, is of course another story). But our freedom as undetermined moral agents is not one whit thereby affected—we are as free to judge incorrectly as to judge correctly; no one can take that away.

The central residual problem is one on which Kovesi, admittedly, says very little; but then neither does Graham. What we will say is regrettably programmatic. To begin with, there are two ways of defining the D-subscript use of predicates. One of these seems to us quite unhelpful; despite the tone of some of his remarks on pages 74 and 78, this is not Graham's definition. One may define a D-subscript use as simply that use of a predicate when a moral agent takes a moral decision about a particular case. This way of defining a D-subscript use makes it merely a matter of time and/or context whether a given use is such a use. Secondly, one may define a D-subscript use as the application of that predicate which sums up the existentially unique total moral worth of that unique act. In this sense, the appropriateness of a given predicate is not merely a function of the fact that at this time this agent applied it. Or, rather, it is a substantial issue in meta-ethics whether it is not. Consider here the difference between Ross and Sartre. Both agree on the impossibility of deducing a unique D-subscript use from a set of G-subscript uses. Ross thinks that an objective intuition comes into play,

and Sartre a subjective creative act. Therefore, contrary to what some philosophers have assumed, the non-deducibility feature or the G-subscript/D-subscript distinction does not entail that moral worth is subjective.

But what alternatives are open to a philosopher who wishes to maintain that nonetheless moral worth is objective? One is classical Intuitionism, but, to cut a long story short, there are problems with that. The most attractive alternative is to apply a theory of reasoning which allows reasoning to be genuine while also doing justice to the uniqueness of particular cases. Insofar as Graham makes his plea for moral freedom without attention to these issues, what he says is seriously defective, though whether one could fairly expect him to say much in the confines of that paper is doubtful. But no reference to the importance of moral freedom in the context of a discussion of moral reasoning is to our mind helpful unless it is set against a scenario such as we have described. It is a fair criticism of *Moral Notions* that Kovesi says very little on this either. In the confines of this paper, we can do little more than gesture. But in our opinion, the emphasis on case-by-case reasoning as a genuine form of reasoning found in the writings of John Wisdom and owing much to the later Wittgenstein lights the way ahead for the exploration of this topic. We do not think the force of Wisdom's theory as applicable to moral philosophy can be appreciated except on the presupposition of a Kovesian view of moral notions. Thus we have considered it important to attempt to put the record straight on what such a view amounts to. Graham's objections are plausible ones, even though we have tried to show they are ultimately ill-founded. A positive account is now required, and this is not the place for that.

University of Alberta

ABSOLUTE PRINCIPLES AND DOUBLE EFFECT

R. A. Duff

I

MORALISTS have traditionally appealed to the Principle of Double Effect (PDE) in order to maintain certain absolute moral prohibitions, such as that against the intentional killing of human beings, in the face of situations which seem to reduce them to absurdity or incoherence. Mr Hanink ('Some Light on Double Effect', ANALYSIS 35.5, 147–151) believes that Mr Geddes' version of the PDE ('On the Intrinsic Wrongness of Killing Innocent People', ANALYSIS 33.3, 93–97) can, when expanded and modified, serve this purpose and be defended against my criticism that it generates sophistical and unacceptable conclusions and reduces both the PDE and the absolute prohibitions it is meant to support to vacuity ('Intentionally Killing the Innocent', ANALYSIS 34.1, 16–19).

On Hanink's account (p. 150), we may perform a single act with both an intended good and a foreseen bad effect, so long as the act itself, apart from its bad effect, is legitimate; the bad effect is not intended, as an end or as a means; and the good effect outweighs the bad. These conditions are satisfied by a craniotomy performed to save the mother's life, and by the setting of an explosive charge under the man who is blocking the only exit from a flooding cave. But in each of the other four cases which I discussed—the cannibalistic killing of the cabin-boy; the use of a man's body for transplant surgery; Gerstein's co-operation with the S.S.; and the execution of a scapegoat—at least one condition is not satisfied: the act itself may be an illegitimate assault, apart from its fatal outcome; the good effect may not outweigh the bad; or the act which has the bad effect is not itself the act which has the justifying good effect but a means to it, so that we cannot outweigh the bad effect by any good effect of the same act.

But his argument leaves two crucial issues fatally unclear, so that his distinctions between the two legitimate and the four illegitimate acts seem arbitrary or inconsistent. He recognises the difficulty of distinguishing between what I must be said to intend, or to do intentionally, and what I may say that I foresee without intending, and allows that there are logical constraints on my descriptions of my intentional actions: that I must include the means by which I achieve my aim (p. 151), and, though rejecting my suggestion that I intend those effects which cannot be intelligibly separated from what I admittedly intend, that 'there is a strong case to be made that a decapitation just is a killing' (p. 150). But why then is not the crushing of the foetus' skull, as an intended means to, or part of, removing it from the mother, or the explosive dismember-

ment of the trapped explorer, as an intended means to clearing the cave-mouth, equally an act of illegitimate killing?

He also recognises the importance, and the difficulty, of identifying and individuating the acts which an agent performs: for he must be able to pick out "the act" which he performs, and say whether the good and bad effects are effects of the same act or of two different acts; but he says only that clarification of these problems must wait until 'philosophers develop more sophisticated theories of act-individuation' (p. 151). Without such clarification, we cannot see why the abortionist's or the explorers' acts are not themselves illegitimate assaults; or why the justifying good effects are effects of those acts themselves, rather than of the further acts which are certainly necessary if the lives are to be saved (caring for the mother, climbing out of the cave). Without clearer criteria for picking out one of the range of possible descriptions of "what he did" as *the* correct description of his act, and for identifying the relevant effects and the acts of which they are effects, we have no good reason for deciding these cases as Hanink advises. His distinctions between them appear arbitrary and unfounded: if we allow any, we have just as much reason to allow them all.

I believe that any plausible absolute prohibition on intentional killing must prohibit all six actions; and that an appeal to the PDE serves, not to distinguish between these actions, allowing some and disallowing others, but to distinguish *within* each case a legitimate refusal to use such means, even to save life, from an illegitimate intentional killing, for however good an end. For in each case the agent can save further lives only by a direct, intentional, and inevitably fatal assault on an innocent victim.[1] However, since I share both Hanink's concern to defend the coherence of absolute moral prohibitions and of the PDE and his realisation of the difficulties of this task, it would be both churlish and unproductive to limit myself to negative criticism. I hope instead to indicate the major misconception which vitiates his attempt, and then to suggest, sketchily and incompletely, a more useful approach: to complete the task would require more space than an article allows, and more insight than I can claim; but a start can be made.

II

Hanink's belief that clarification of the distinctions between intended and foreseen effects, between means and side-effects and between one act and another must await the development of more sophisticated theories of act-individuation suggests that these distinctions and the criteria for drawing them must be established prior to, and independently

[1] Compare the comments in Smith's case (D.P.P. v. Smith; (1961) A.C.290, at 327) on 'unlawfully and voluntarily doing something to someone'. My later comments will show that I am not here trying to provide a general descriptive criterion for the identification of intentional killings.

of, any moral interest in them; that we must first establish clear-cut and general criteria, and then raise the question of their moral significance. This is a common view: it underlies Bennett's attempt to give a general account of the distinction between acting and refraining which is detached both from the particular contexts in which such a distinction is invoked and from the moral perspectives which require it.[1] It fits with a more general view of the status of moral principles and their relationship to particular moral judgements which is most clearly articulated by Hare:[2] particular judgements are justified by showing them to be the conclusions of practical syllogisms whose major premisses are universal moral principles, specifying clear and descriptive criteria for the application of moral terms of commendation and condemnation. Thus an absolute prohibition on the intentional killing of human beings, from which our particular judgements are derived and by which they are justified, must specify criteria for the identification of such actions: the PDE is a more general moral principle, which provides such criteria and asserts that we may sometimes bring about, with foresight but without intent, an effect which we may not bring about intentionally. The criteria it provides for distinguishing intended from merely foreseen effects must (given the fact-value distinction) be descriptive, or morally neutral, and universal: we must be able to formulate and apply them in advance of any moral view of their significance, and *then* go on to discuss their moral relevance; they must be valid for, and equally relevant in, all cases in which the distinction can be drawn.

This is a fundamentally mistaken view of the logical grammar of moral principles, and thus of the proper approach to an understanding of absolute moral prohibitions and the PDE. Bennett's own account of the distinction between acting and refraining in terms of the number of alternative bodily movements by the agent which would have been followed by the same effect shows, not that the distinction lacks any intelligible moral significance (although *his* distinction certainly does), but that this is a quite inappropriate way of analysing the distinctions we draw between acts and omissions, refrainings, failures to act, and so on.[3] In abstracting the distinctions from the (often moral) contexts in which they have their meaning, he has abstracted from them just those features which are crucial to understanding them: he is like someone who, seeking to understand the emotional and moral significance of sex, begins with a general account of sexual activity in purely physiological terms and is then puzzled by our attaching importance to this kind of activity.

I have space here only to assert, not to argue for, an alternative view

[1] J. Bennett, 'Whatever the Consequences', ANALYSIS 26.3, 83–102.
[2] R. M. Hare, *The Language of Morals* (OUP 1952); *Freedom and Reason* (OUP 1963).
[3] Cf. J. Casey, 'Actions and Consequences', in J. Casey (ed), *Morality and Moral Reasoning* (Methuen 1971); D. Dinello, 'Killing and Letting Die', ANALYSIS 31.3, 1971, 83–6.

of the logic of moral principles and of the criteria for identifying and distinguishing actions which they involve.[1] Three points must be emphasised. First, moral distinctions such as that between intention and foresight are meaningful only within particular contexts, viewed from particular moral perspectives: in explicating them, we are not trying to establish universal principles, specifying descriptive criteria for their application, but to show how, in those contexts, within those perspectives, they can significantly be drawn. Any general principles we propose serve, not as the major premisses from which we may then derive particular judgements, but rather as slogans or reminders of what we have established in the particular cases: they need not, and cannot, themselves provide clear criteria for deciding future cases.

Secondly, although we must obviously show that the judgements we make and the distinctions we draw are consistent between different cases, this does not involve showing that they all fall under the criteria provided by some universal principle. The criteria, both for judgements and distinctions and of consistency, are provided *within* a moral perspective: they are displayed, not by specifying universal principles, but by showing in the discussion of particular cases the relevance and significance of identifiable resemblances and differences. But, thirdly, we cannot hope to specify the relevant distinctions, resemblances and differences in purely descriptive, morally neutral, terms: the identification of relevant act-descriptions, the distinctions between intentional acts and foreseen side-effects or omissions can be made only within a moral perspective and the values it involves. "Pure" descriptions can provide no basis for moral judgement or understanding: the morally relevant descriptions and distinctions are already morally loaded, in that they are possible and meaningful only within a moral perspective.

This is not to say that our description of a particular case is determined by a prior moral judgement on it: an action is right or wrong *because* of the kind of action it is, *because* of the description we can or must give it. Our descriptions are determined, not by our particular judgements, but by our general perspective on human action, and on human rights, roles and duties. Thus our account of the craniotomy case—our view as to whether the doctor would be intentionally killing the foetus if he operated or the mother if he refused to operate—will depend on, or be part of, our general view of the extent and nature of a doctor's duties to, and rights over, his patients: and this will be related to our views about the rights and duties of anyone with regard to the lives of others. We justify the judgement that the operation would involve an intentional and illegitimate killing by showing how it constitutes a direct and necessarily fatal assault on the foetus; and that a refusal to operate

[1] For a fuller account of the view I am here relying on, cf. D. Z. Phillips and H. O. Mounce, *Moral Practices* (RKP 1970); R. W. Beardsmore, *Moral Reasoning* (RKP 1969).

does not amount to wilfully killing the mother by showing how, although a doctor's duty to care for his patients is such that a refusal to do what he knows to be necessary for their survival would in most cases constitute a wilful killing, he may not contemplate killing as a means to saving life, and may thus deny responsibility for a death which could only be averted by killing someone else.[1] We must make sense, within the context of that situation and of more general moral considerations, of these descriptions and distinctions: but this does not involve bringing the doctor's action under some universal principle specifying descriptive criteria for the identification and description of his action.

My suggestion is that we can understand absolute moral prohibitions and the PDE only through a detailed examination of the moral perspectives within which they are invoked and of the descriptions and assessments of particular cases which these perspectives involve. This approach can be clarified and justified only by showing how in practice it does enable us to understand more adequately the moral values and distinctions with which we are concerned. But I must first comment briefly on one common, but misconceived, criticism of such an approach to the understanding of moral or religious beliefs: that this talk of moral perspectives, or of language-games and forms of life, saves the beliefs in question from external criticism only at the cost of imprisoning them within self-contained systems which preclude any argument, or even understanding, between different perspectives.[2] This criticism mistakenly assumes that each moral perspective is meant to be circumscribed by clearly-defined boundaries which separate it absolutely from any other perspective, as a self-contained and self-sufficient system of thought and experience. But this is not the case: we belong to the same society and share a common language, *within* which we find a variety of perspectives; which is to say that these perspectives are logically connected; that they overlap, connect with and run into each other. An Absolutist's view of the definition and the ethical significance of killing and of distinctions such as those between intended and merely foreseen effects is different from more ordinary views: but his views are related to ours, and an attempt to understand them is not just an attempt to show the internal consistency of a self-contained system, but rather to show, by tracing out the connections between his conceptions and our own, how this is an intelligible perspective for a man to hold. This will involve both showing how the actions and distinctions to which he attaches significance are closely related to those which we regard as significant, and also appealing for an imaginative understanding of a perspective in crucial respects different from ours.

[1] Cf. Casey, op. cit.

[2] Cf. A. MacIntyre, 'Is Understanding Religion Compatible with Believing?', in J. Hick (ed), *Faith and the Philosophers* (Macmillan 1964).

III

An understanding of the PDE must begin with the contexts in which it is invoked. Philosophical discussions sometimes portray it as an ad hoc device, designed to save an absolute prohibition on, for instance, the killing of innocent human beings from the difficulties presented by cases in which, through no fault of his own, an agent cannot ensure that no one dies, but only determine by his actions who and how many should die. For if we allow that he kills anyone whose death he could have averted he will break the absolute prohibition whatever he does: but it is intolerable that I should, through no fault of mine, be unable to avoid committing an absolute crime; and absurd that our moral principles, whose function is to guide our actions by showing us the right thing to do, should fail to provide such guidance here. So the PDE is invoked to show that at least one of the agent's options would not involve an intentional killing and that he can thus act rightly and without guilt.

This gives a false impression of the significance of the PDE: indeed, an Absolutist could allow that we may sometimes be unable to avoid an absolute crime, and deny that the primary function of moral principles is to enable us to pick out *the* right action in any situation.[1] But the distinctions summarized in the PDE are already intrinsic to his moral perspective on human action—as they are to our own: for he is concerned to prohibit certain *actions*, not just certain events, and thus already has some conception of intentional human action. This concern with intentional agency is the first important feature of an Absolutist perspective which must be emphasised: it can best be clarified by contrast with the strictly Utilitarian view of agency to which it is opposed.

A Utilitarian regards outcomes or events, rather than actions, as of primary moral significance. His concern is to prevent or assist the occurrence of certain kinds of event which are described and assessed in terms of their effects on human happiness or well-being: he is interested in human agents and their actions only secondarily, insofar as they can have an effect on what happens and insofar as we can bring them to effect the consequences we want; it is as producer or cause of relevant consequences that the human agent is significant. Thus an agent's responsibility for a certain outcome—his duty to take it into consideration in his practical deliberations, his answerability for its occurrence or non-occurrence—depends on the utility of holding him responsible for it: and this is determined primarily by the extent to which he does or can foresee and control it. He is the agent of, and responsible for, all those effects which he does or can foresee and control. Distinctions between intention and foresight or between acting and refraining can thus have no intrinsic moral significance: they are morally relevant only insofar as they are correlated with relevant differences in the utility of holding

[1] Cf. Phillips and Mounce, op.cit., ch. 8.

people responsible for consequences to which they are related in these different ways; with different degrees of foresight or control; with different kinds of motive, having different degrees of general utility; with different degrees of freedom which our moral demands would leave to an agent.[1] To attribute any intrinsic moral weight to them is, at best, to mistake pragmatic rules of thumb for absolute principles; at worst, it is an attempt to evade our proper moral responsibility for all those effects which we can, by our actions, control.

The Absolutist, on the other hand, is primarily concerned with the intentional actions of human agents, rather than their consequences. What matters is not simply that an event occurs which I did, or could, foresee and control, but the way in which I am related, as an agent, to that event: what matters is what I *do*; and "what I do" is determined not just by what happens, but by the intention revealed in my action and by the complex context of relationships, responsibilities, rights and duties within which I act. His absolute prohibition is against the intentional action of killing, not against the occurrence or the foreseen and avoidable causation of death: it would indeed be absurd to prohibit *that* absolutely, since for any prohibited outcome we could imagine a case in which the outcome of any alternative is even worse. His first task is thus to give a clearer account of this notion of agency: to show the sense that it has, in particular contexts, and to show how it is part of an intelligible moral perspective of human life.

This task is simplified by the fact that we ourselves attribute a similar significance to agency. We draw moral distinctions between what we intentionally do and what our actions foreseeably cause or what we fail to prevent; between the harm we intentionally cause and that which we fail to avert or which occurs as the by-product of some other intentional action; between trying to harm someone and recognising that he will or may be injured by what we do: and we may sometimes allow the one while condemning the other. Such distinctions are part of our ordinary moral life. This is not to say that we are all Absolutists, or that an Absolutist differs from us only in prohibiting absolutely actions which we might sometimes allow: for his account of these actions and distinctions, and their significance, will still be crucially different from that of others who still allow them some significance. But it reminds us that his views are not entirely alien to ours, and that in explaining them we will partly be explaining our own and showing how his are an intelligible modification of them. It reminds us too of the extent to which a strictly Utilitarian view is at odds with our ordinary moral attitudes; but to make sense of non-Utilitarian conceptions of agency constitutes neither a disproof of Utilitarianism nor a proof of Absolutism. Our philosophical

[1] Cf. P. J. Fitzgerald, 'Acting and Refraining', ANALYSIS 27.4, 133–9.

concern is with the more modest, but vital, task of explicating the meaning and intelligibility of an Absolutist perspective.

The second feature of an Absolutist perspective which must be emphasised is its concern with *transcendent* values—with moral demands which transcend all purely human interests. In one sense we are all Absolutists; any moral view must recognise *something* as having an absolute value: for a Utilitarian only the increase of human happiness is absolutely valuable, while for those more usually called Absolutists certain actions have an absolute value or disvalue independent of their contribution to human happiness. But there is no room for an Absolutist within a secular perspective which bases all values on some conception of human well-being. Even if this could include certain intrinsic, non-Utilitarian, values, internally related to a suitably rich conception of human well-being, it cannot *absolutely* require or prohibit any particular actions; for we will always be able to imagine a situation in which the alternative is more conducive to, or more destructive of, human well-being. An Absolutist's moral demands must have a status and a force independent of, and prior to, any conception of human well-being. Only in the light of an allegiance to a transcendent, "other-worldly" Good, which imposes its demands on us and by reference to which we must determine and assess our conceptions of human well-being, of human interests and their importance, can we see certain actions as absolutely right or wrong.[1]

Absolutist ethics, like those of Kierkegaard, Simone Weil and of traditional Roman Catholicism, are often set in a specifically religious context. But a recognition of absolute and transcendent values need not depend on belief in a God whose Will they express: some of Camus' writings, for instance, reveal an attempt to provide a nontheistic account of certain absolute moral demands and constraints on human action. Each different kind of account will need detailed examination: I wish here only to emphasise the crucial role of transcendent values in any Absolutist perspective and the way in which this helps to clarify the Absolutist's concern with agency. For he is concerned, not with the production or prevention of outcomes in this world, but with our relationships as agents to the Good, with our activities as willing, intentional and purposive agents in relation to the moral demands of transcendent values. Since these values are not of this world, they may require us to act in ways which are, by the standards of this world, disastrous: to refuse to sanction an abortion even when this is the only way to save the mother's life, and when both mother and foetus would otherwise die; for the evil of an intentional killing does not just outweigh

[1] I have tried to say something more about the relationship between such a conception of a transcendent Good and human well-being in 'Must a Good Man be Invulnerable?', forthcoming in *Ethics*.

(which would sugggest commensurability), but is of a different category altogether from, the evil involved in the death of mother and foetus.

This brings us to a third feature of an Absolutist perspective, its concern with *limits*. It is no accident that absolute moral demands are usually expressed as prohibitions; for much of their significance lies in the limits they set on our rights and responsibilities as moral agents. This point is missed if we take the Absolutist to be saying that certain actions, like intentional killing, are always worse than any alternative; for this suggests that he has weighed the commensurable evils of killing and of not killing in a range of cases and is now committed to a prediction about their relative weights in *all* possible or conceivable cases. But such an inductive prediction cannot justify an *absolute* prohibition: we cannot rule out a priori the possibility or conceivability of a case in which the balance will come down on the other side, in which the evil results of not killing will be so massive that they require us to kill as the lesser of two evils.[1] The Absolutist is not concerned with the relative and commensurable evils of alternative courses of action: intentional killing is not an alternative, whose merits and demerits must then be weighed, but something which we may not even contemplate as a possibility. Once we realise that a course of action would involve an intentional killing, it is closed to us, however good its effects might be.

He takes a more humble and modest view than a Utilitarian of our status as moral agents. For a Utilitarian, *any* action could in principle be justified as an essential means to the maximisation of human happiness. The constraints and prohibitions which we recognise are provisionally and pragmatically justified by the probability of disastrous consequences flowing from the actions they preclude and by the dangers, from human fallibility and bias, of allowing us to contemplate such actions. But the Absolutist accepts certain *absolute* constraints even on his attempts to benefit others. His responsibility to save life and his rights over the lives of others in doing so are strictly limited by the requirement that he should not contemplate intentional killing. If he could save life only by killing, he is not guilty of the deaths he fails to prevent; for they are no longer his responsibility.[2] We can understand this notion of absolute limits on human action only if we remember the Absolutist's concern with our relationship, as moral agents, to a transcendent Good.

I have tried in this section to outline some of the features essential to an understanding of an Absolutist perspective. This is clearly only a beginning, but both critics and defenders of Absolutist views have too often begun with a misconception of what they are attacking or

[1] Cf. Hare, *Freedom and Reason*, 42–4, for this kind of criticism of Absolutism.

[2] Cf. G. E. M Anscombe, 'Modern Moral Philosophy', *Philosophy*, 33, 1958, 11–12, on not being responsible for the bad consequences of one's good actions.

defending. Hanink's comment on the abortion case, that 'If we suppose that failing a craniotomy both mother and infant will die and that allowing this is folly, it is essential to construe the operation as Geddes does (i.e. as not involving an intentional killing)' (p. 148), reveals such a misconception. He takes it as obvious that no acceptable principle could require us to leave two people to die when we could save one by hastening the death of the other. This is an intelligible moral position: it allows the killing, not of anyone whose death would enable us to save other lives, but only of those who would anyway have died, and perhaps only if they would have died from those causes which would also have caused the deaths we hope to avert.[1] Given a fuller account of the position and of the cases involved, it could also allow the explorers' action and some or all of the actions which Hanink disallows; but what it allows is the intentional killing of a human being. It is not the position of one who believes that it is *absolutely* wrong to kill; for to take that view is to see one's actions and obligations in the light of moral demands which transcend all purely human interests and in obedience to which we may never contemplate killing, however "foolish" in practical or Utilitarian terms this might be.

IV

I have argued that an adequate explication of Absolutist moral perspectives requires a long and detailed examination of the contexts in which absolute moral demands are invoked and of the sense we can attach, in such contexts, to the crucial notions of transcendence, agency and limits on human action. Some brief comments on the case of suicide, which Hanink mentions at the end of his paper, may indicate how such an examination should proceed.[2] We may condemn suicide, or some suicides, for a variety of reasons: by reference to the harm done to others, or to the agent himself; by suggesting that suicide marks a lack of courage in facing up to life's difficulties; by arguing that a man owes his life to the state and may not dispose of it as he wishes. But none of these considerations could generate an *absolute* prohibition on suicide; for we could always find a case in which they themselves justified suicide, in which the avoidance of misery for oneself or others or the demands of courage or the interests of the state required a man to kill himself. An absolute prohibition on suicide must rest on a transcendent moral value, which denies to anyone the right thus to dispose of his own life. This makes most obvious sense in the context of a religious belief, that my life is God's and must be left in his hands, that it is not up to me to

[1] Cf. B. A. Brody, 'Abortion and the Sanctity of Human Life', *American Philosophical Quarterly*, 10.2, April 1973, 133–40, for the development of this kind of position.

[2] On an Absolutist view, and definition, of suicide, cf. especially R. F. Holland, 'Suicide', *Royal Institute of Philosophy Lectures, II, Talk of God*, G. N. Vesey (ed), (Macmillan 1969).

decide when it should end: this is one of the limits set on our rights and responsibilities as moral agents.

But this is not to say that it can never be right to act in a way which I know will result in my own death; for not all such actions are suicides or acts of *intentional* self-destruction. How then can we distinguish suicide from, for instance, heroic self-sacrifice, and show that the latter need not involve intentionally killing myself or improperly arrogating to myself the responsibility for deciding when my life should end? Our ordinary conception of suicide embodies some such distinction: we would not ordinarily count as a suicide a man who resists an unjust regime or who stays in a blazing house to save others, even when he knows that he will die as a result. But the Absolutist's distinctions and the significance he attaches to them will be in crucial respects different from, and more refined than, those which we recognise. Both our definitions of, and our moral judgements on, suicide will differ; and a definition of suicide can be spelled out only within the moral perspective from which suicide is viewed.

Captain Oates, for instance, walked out to certain death in a blizzard to give his friends a better chance of survival. A non-Absolutist might call this a heroic suicide, in view of the certainty of his death; or, to emphasise the extent to which Oates' attention was fixed on saving others, he might deny that it is "really" a suicide and talk rather of heroic self-sacrifice. The flexibility of his moral perspective on suicide allows an equal flexibility in the definition of the term. And he would be unlikely to attach any intrinsic significance to the difference betweeen walking out into the blizzard and, for instance, shooting himself. But this is a crucial difference for an Absolutist, who would deny that walking into the blizzard is an act of suicide, but insist that shooting oneself is, and is thus an absolute crime, even given the same admirable motive of saving one's friends. Death, we may suppose, is equally certain in either case, the end aimed at—of bringing his friends to go on without him, which he knows they will not do while he is there with them—is the same; but the means adopted are crucially different. For in one case they will go on because he is dead, and he intentionally kills himself, by shooting, as a means to this. But in the other case, if the Absolutist justification is to be possible, he intends them to go on because they realise that he has chosen to withdraw from the group; and to achieve this, he needs simply to walk away.

Of course, he knows, and they know, that he will certainly die; but this is now a consequence, not a part, of his intentional action. It is separable from it, in a way in which his death is not separable from shooting himself. This separation—this logical gap between what he intentionally does and his consequent death—is important, not because it allows him or them to hope that he will in fact survive (they had no such hope), but because it shows that his intentions, and attention, need in no

way be directed towards his death. All he is deciding to do is to walk away: the rest is up to God. To say this, or to say, as he walks out, 'My life is in God's hands', does not express a hope that God may save him from death: it is rather to say that that aspect of his action—its implications for his own life or death—is and should be no part of his present concern; and this is possible only so long as his death is not an essential means to the end he is pursuing.

The distinction between suicide and heroic self-sacrifice is reasonably clear in Oates' case: but it becomes more problematic in cases where it is harder to separate the death from what is intentionally done, as in Hanink's case of the man who throws himself on a live grenade to save his friends (p. 151). Hanink argues that this is a counter-example to my suggestion that I intend those effects which cannot be intelligibly separated from what I admittedly intend, since we cannot intelligibly imagine a man surviving this, but should still not call it a case of intentional self-destruction. I do not believe that this case is a counter-example, since we need only imagine the grenade not exploding to imagine the man surviving: what he intentionally does is to interpose himself between his friends and a serious danger, which is no more a suicide than Oates' action. However, I am still worried by cases like this, partly because of the difficulty of drawing the necessary distinctions in each case, and partly because we need to explain the asymmetry between cases involving the agent's own death and cases involving the deaths of others; for while the Absolutist may deny that Hanink's hero is a suicide, he would presumably say that someone who threw someone else on to the grenade was guilty of murder—of intentionally killing him. Yet the victim's death is surely equally separable, or inseparable, from what is intentionally done in either case.

I have no adequate answer to offer to these problems yet, save to emphasise again that they can be clarified and resolved only by a thorough examination of the contexts in which they arise and the moral perspective within which they are problems. My suggestion certainly does not provide, and was not intended to provide, a complete or adequate criterion for deciding what is to be included in a description of a man's intentional action. For a crucial role in determining the logical and moral constraints on such descriptions is played by the moral perspectives within which the actions are both described and assessed: these determine the duties which we have with regard to our own lives and those of others, and thus the aspects of our actions, insofar as they affect these lives, to which we have a duty to attend and of which we are accordingly held to be the intentional agents. An Absolutist might thus explain the asymmetry between suicide and murder by arguing that our duty with regard to our own lives allows us sometimes to ignore, to exclude from our attention and thus from our description of what we are

intentionally doing, the fact that our action will have fatal consequences for ourselves, while we could not thus ignore and exclude it if the consequences were fatal for someone else, even when in all other respects the cases were symmetrical. But we can clarify, and assess, this suggestion and the other problems only in the course of a far longer and more detailed study than I have space for here.

University of Stirling

'THIS IS FALSE' ON THE PROSENTENTIAL THEORY

By DOROTHY L. GROVER

I

WHAT account can be given, within the prosentential theory of truth, of the liar sentence 'This is false'? There can be no suggestion of ascribing to it (or any proposition it might be said to express) the contradictory properties truth and falsity, since, according to the prosentential theory, there are no such properties. The intuition that we have that there is something wrong with the liar sentence can be accounted for by the theory, however, and so also can the intuition that it is not all bad.

In discussing these points I appeal to some claims concerning proforms of laziness made by Joseph L. Camp, Jr., Nuel D. Belnap, Jr., and myself in 'A prosentential theory of truth' (*Philosophical Studies*, Vol. 27, 1975, pp. 73–125).

(1) Proforms and modified proforms, upon given occasions of their use, get their semantic content from their antecedents. (This is not quite accurate, but it is good enough for the points I wish to make.) Thus, in the easy cases, an anaphoric 'she' and its antecedent are co-referential; in the case of a prosentence, it is the propositional content of the antecedent which is shared by the proform. For example, in

> Trudy: Robert can be helpful.
> Mavis: That is true.

Mavis uses 'That is true' to assert, with anaphoric overtones, what Trudy asserted. A modified prosentence has as its propositional content a proposition which is the result of applying an appropriate operation to the proposition expressed by the antecedent. The semantic content of a proform of laziness thereby changes from context to context:

in the case of a pronoun, its referent changes, and in the case of a prosentence, the proposition it expresses changes.

These semantic facts concerning proforms of laziness are reflected in the following syntactic facts.

(2) (Again oversimplifying.) The anaphoric substituend of a proform, if there is one, is (possibly a grammatical variant of) its antecedent, and in the case where the proform is modified, its anaphoric substituend is an appropriately modified form of the antecedent. For example, the anaphoric substituend of the modified prosentence 'That might be true' in

Judy: There are people on Mars.
Bob: That might be true.

is 'There might be people on Mars'.

Accordingly, proforms of laziness are said to "stand in" for their antecedents.

An immediate consequence of all this is that proforms of laziness tend to inherit the semantic problems of their antecedents—they do not have lives of their own. Consider

Bill: The King of France is wise.
Jane: And he is also bald.

Since 'The King of France' is the antecedent of Jane's 'he', any failure by Bill successfully to refer when using 'The King of France' brings about a similar failure in Jane's use of the anaphoric 'he'. Similarly for prosentences. Consider

Bill: The King of France is wise.
Jane: That is true.

'That is true' inherits the problems of its antecedent: if, because of the reference problems, one has a theory which construes 'The King of France is wise' as expressing only a quasi-proposition, then so will 'That is true' in this context have only quasi-propositional content.

Now to the liar sentence. 'This is false' is a modified form of the prosentence 'This is true'. Standing on its own, as it so often does, 'This is false' looks like a proform of laziness without an antecedent. Since proforms of laziness get their semantic content from their antecedents, it might therewith be concluded that 'This is false' fails to have semantic content: just as, if 'He is tall' is uttered without an antecedent or other semantic determining context, 'he' fails to have a referent, so also, it might be argued, 'This is false' fails to express a proposition. I agree with the conclusion, but not for that reason: 'He is tall' does not provide the best analogy, since 'This is false' does have an antecedent—itself.

A better analogy is one where an anaphoric pronoun has as its antecedent a noun phrase which in its turn purports to pick up a referent

from the pronoun. Such a case exemplifies the semantic circularity found also in the liar sentence. For example

> Sam: The person you are going to refer to in your next sentence happens to be a fool.
> Sally: He is not.

Despite the fact that the anaphoric 'he' has an antecedent, neither speaker succeeds in referring. The reason is that neither the antecedent nor the pronoun, either on its own or in combination with the other, fixes a referent. Similarly, *the liar sentence lacks propositional content*—neither as an antecedent nor as a modified prosentence can 'This is false' introduce propositional content. It is interesting to note that 'This is true' is no better off.

At this point, and as a check on the above, one might consider looking for an anaphoric substituend for 'This is false'. The anaphoric substituend must be a sentence which can be used to assert the contradictory of its antecedent 'This is false'. Because of the indexical (embedded in a prosentence) in the antecedent, the "obvious" candidate 'It is false that this is false' won't do. More appropriate is something like

This is false. That is false.

Since prosentences cannot, on their own, introduce propositional content, this doesn't help: furthermore, since whatever we try will always consist of prosentences and modified prosentences, no anaphoric substituend will discover a proposition for 'This is false' to express.

What 'This is false' has going for it is the fact that *the grammatical picture is complete*—it is a proform *with* an antecedent. This is in contrast to a 'That is false' standing alone. This fact explains why, on the one hand, one is momentarily tempted to puzzle over its sense, and yet, on the other, not tempted to puzzle over an isolated 'That is false'.

While an account which construes the liar sentence as failing to have propositional content is not new, there are two respects in which the explanation provided by the prosentential theory is simultaneously both attractive and not ad hoc: as Susan Haack has pointed out in correspondence, it does not rule out all self reference; secondly, this prosentential explanation of the semantic facts flows naturally from a theory of truth which was introduced to explain other, more important, features of our truth talk.

II

There are also quantificational versions of the liar sentence. For example

(Q) Anything which is identical with this sentence is false.

Since no one has yet provided an adequate theory of quantification in English, it is not possible to explore in detail the ramifications of the prosentential theory with respect to (Q). If it should turn out, however, that the propositional content of (Q) is determined by its substitutional instances, then (Q) may, like 'This is false', have no propositional content.

If the substitutional instances are defined in the "usual" way, then some version of the following will, I assume, be a substitutional instance of (Q).

> If 'Anything which is identical with this sentence is false' is identical with 'Anything which is identical with this sentence is false' then it is false.

The antecedent of the modified prosentence 'it is false' is the quoted sentence together with the quotes. (In case it seems strange that the antecedent of a prosentence is sometimes a noun, I remind the reader of the following: anaphoric pronouns do not always have nouns as their antecedents, and sometimes, as in the following, the antecedent is a sentence.

> John visited us. It was a surprise.)

'It is false' gets its propositional content from its antecedent. But what proposition does (Q) express? It is at this point that a problem arises since the antecedent does not *yet* have semantic content: as in the case of 'This is false', neither the antecedent, nor the proform, introduce the required semantic content.

University of Illinois at Chicago Circle

MERCY AND FORGIVENESS

By P. TWAMBLEY

MERCY and forgiveness are qualities with important similarities, but these features have not always been correctly identified. The accounts of mercy and forgiveness I will be contesting show a significant parallelism. Underlying both there is, as I will try to show, a view of morality which threatens our understanding, not only of mercy and forgiveness, but of generosity, kindness and the other miscalled 'duties of imperfect obligation'. In place of these incorrect accounts I will propose a positive account which I hope will bring out the genuine parallels between mercy and forgiveness. My discussion of mercy is indebted to a discussion-note by H. R. T. Roberts 'Mercy' (*Philosophy*, 1971).

The view of mercy I am concerned to contest is that of Alwynne Smart ('Mercy', *Philosophy*, 1968) and of forgiveness that of the late Aurel Kolnai ('Forgiveness', *P.A.S.*, 1973–4). The latter's position is rather more complex than my discussion may suggest. But the emphasis is certainly as I describe and, in this case, it is the emphasis which I think it is important to readjust.

I

'A theory of punishment should give some account of mercy' is Mrs. Smart's opening contention, and throughout her discussion she construes mercy as a quality the exercise of which is peculiar to a *judge*, one who has authority to punish. Thus, the question she sets herself to answer is 'What are the conditions for the appropriate exercising of mercy, how do we decide how much mercy is appropriate, and when is a judge morally obliged to be merciful, if ever?' (p. 345). In seeking the answer to this question, Mrs. Smart concentrates exclusively on examples drawn from the criminal courts. The justice mercy is to temper is taken to be criminal justice.

In this first part of her paper Mrs. Smart correctly puts on one side such things as a jury's recommendation of mercy and pleas for mercy on the grounds of mitigating circumstances. She concludes that such cases of "mercy" are simply misnamed, for here 'mercy is nothing more than a way of ensuring that a just penalty is imposed and injustice avoided' (p. 349). In these cases "mercy" simply collapses into justice.

In the second section Mrs. Smart discusses what she believes to be genuine but inappropriate exercises of mercy. What then is a genuine case of mercy? 'When a man exercises mercy what he does is acknowledge that an offence has been committed, decide that a particular punishment would be appropriate or just, and then decide to exact a punishment of lesser severity than the appropriate or just one' (p. 350). The main

point of exercising mercy is to avoid suffering. Given this account it is scarcely surprising that the occasions on which mercy is justified are severely limited. Mercy is unjustified 'if it causes the sufferings of an innocent party, is detrimental to the offender's welfare, harms the authority of the law, or where it is clear that the offender is not repentant or not likely to reform' (p. 350). Further, a judge may be rightly criticized for showing favouritism and committing an injustice if he shows mercy to one man but not to another whose offence and circumstances are identical in the relevant respects. Thus it seems that at one extreme mercy collapses into justice, at the other its exercise is unjust.

However, Mrs. Smart does think that there are justified cases of mercy. The paradigm is where a judge shows mercy, not because the offender deserves it, but because it is necessary if he is to meet the claims that other duties have on him. Thus he may reduce an appropriate fine because it would impose an intolerable burden on the offender's wife and family. Mrs. Smart concludes 'If we regard mercy as deciding, solely through benevolence, to impose less than the deserved punishment on an offender, then the answer to the original question "when are we justified in being merciful?" must be: only when we are compelled by the claim that other obligations have on us' (pp. 358–9). It is a corollary of this (according to Mrs. Smart) that mercy is intelligible only within a multi-principle conception of morality.

I disagree strongly with this account. Mercy, properly understood, has no essential connection with punishment. It is not the prerogative of a judge: on the contrary a judge has no right to be merciful. In accepting his office a judge places himself under an obligation to impose just sentences and to treat like cases alike.

What Mrs. Smart has in fact discussed is best characterized as *leniency*. The question of leniency does arise with respect to punishment, and Mrs. Smart has accurately located some of the difficulties to which it gives rise. In general, I think it is true to say that, though a judge is under obligation to deal justly, his office carries with it a degree of discretionary power to be lenient in cases of hardship such as Mrs. Smart describes. The reverse side of this coin is seen when a judge imposes a harsh sentence as a deterrent to would-be offenders.

II

Mrs. Smart was not altogether mistaken in going to court to understand mercy, but I suggest she has gone to the wrong court. Not the criminal court, but the *civil* court, is where we will find an adequate model for mercy. The model I propose takes as its centre-piece a plaintiff having a right over a defendant. The exercise of mercy consists in the plaintiff waiving his right over the defendant and thus releasing him from his "bond". I have classical authority for this model. *The Merchant*

of Venice contains probably the most celebrated eulogy of mercy in the language, and in this it is Shylock, the plaintiff (*not* the judge), whom Portia bids temper justice with mercy. Shylock is *within his rights* to demand his pound of flesh, it is his 'in justice'. Portia begs him to waive that right, to exercise mercy. Thus we see that the justice which mercy tempers is not criminal, but what we might call 'civil', justice.

In contrast to this explicitly legal case, my second example of mercy is taken from Joseph Conrad's magnificent story, 'The Duel'.[1] General Feraud and General d'Hubert have been at odds for years, duel has followed inconclusive duel. As Conrad writes, 'For years General d'Hubert had been exasperated and humiliated by an atrocious absurdity imposed upon him by this man's savage caprice'. It has been a blight on d'Hubert's life. In their final duel Feraud fires twice and misses, d'Hubert has his two shots remaining. By every rule of single combat Feraud's life belongs to him. But, despite all the injury done to him, d'Hubert waives his right, he leaves Feraud unharmed. The true nature of his act is brought out by the great delicacy of his behaviour: 'In anger he could have killed that man, but in cold blood he recoiled from humiliating by a show of generosity this unreasonable being . . .' In this concern not to make a show of his mercy d'Hubert illustrates the gratuitous nature of mercy. His manner of acting, his whole attitude, is totally different from that of a judge in being lenient to an offender for his family's sake. And d'Hubert's action does not end there. He goes on to support Feraud anonymously, to keep him from penury. He observes to his wife 'I had the right to blow his brains out; but as I didn't, we can't let him starve'.

To generalize, one man shows mercy to another when he waives his right over that person and thus releases him from his obligation, cancels the debt.

Keeping this in mind, we may see more clearly the following points:

(i) There is no *essential* relationship between mercy and punishment. One who exacts a debt does not *punish* his debtor: even d'Hubert would not have punished Feraud had he killed him.

(ii) Related to this is the fact that a man who is shown mercy may not have actively *offended* the person by whom he is shown mercy. It is enough that he *owes* him some form of debt, that the merciful man has a right over him.

(iii) There is no question of mercy being unjust nor, I would contend, is it ever unjust not to show mercy. It would only blur useful distinctions to say that Shylock was unjust in demanding his pound of flesh. He was hard, cruel, barbaric, certainly, but he was also "within his rights".

[1] I am grateful to Professor R. Holland for drawing my attention to this story; he discusses it at length and with great sensitivity in an unpublished paper.

(iv) Similarly, to talk of a man being *under an obligation* to be merciful is in most cases misleading. (An exception is where the plaintiff has made a promise to release the defendant from his bond.) Again it might show him to be a wicked, despicable, utterly immoral man if he does insist on his rights, but talk of 'obligation' tends only to obscure this. Talk of 'obligation' seems most in place where performance or abstinence can be demanded *by right*.

(v) Mercy is not in tension with treating like cases alike, the merciful man is no more obliged to do so than is the generous man.

(vi) All this, I hope, brings out a central feature of mercy: (in normal circumstances) mercy is *freely* given, it is not something to which a man is "compelled" by the claim of other obligations, as Mrs. Smart would have us understand (cf. pp. 358–9). Rather, as Shakespeare put it, 'The quality of mercy is not strain'd'.

It remains to clear away some apparent anomalies. It may be objected to this account that people ask for mercy from brigands, gunmen, assailants, etc., people who have no rights over them. This is true, though I think my account renders this practice intelligible, for there is a close link between having rights over someone and having *power* over them. But—and this brings out the difference—supposing the gunman accedes to the plea, has he been merciful? Surely not, for he had no *right* to injure his proposed victim. If the gunman had ignored the plea, we might well have called him 'merciless', but 'merciful' and 'merciless' are not true opposites. One who is not given to the exercise of mercy may be harsh but need not be altogether merciless. I think Mrs. Smart is correct in suggesting that now 'merciless' is virtually synonymous with 'cruel'.

The Merchant of Venice might seem to raise a more substantial difficulty. For, after Shylock fails to show mercy to Antonio, not only does Antonio show mercy to him by relinquishing his claim on Shylock's property, but the Duke does also by sparing his life. Yet the Duke is a judge. It might be possible to construe this as an act of leniency, but I am not sure that this would be correct. It is, I think, an act of clemency. But even as such it can be accommodated. For the Duke exercises mercy *as* the head of the Republic: by attempting the life of a member of the Republic, Shylock's life is the State's by right. The Duke (and he only) has the right to waive that claim. But note: the Duke has power to do so *not* as a judge, but as head of State. Similarly, in Britain the Home Secretary has the power to grant a pardon, to exercise clemency, not as a private individual, nor as a judge, but in his (fictitious) office as right-holder. (To repeat: judges have no right to be merciful because it is not *to them* that any obligation is due. And they have an obligation to impose the sentence the law prescribes.)

In discussion Stewart McNeilly observed that ours is a 'merciless society'. By this he drew attention to the slight absurdity of asking for mercy where (say) a small debt was in question. 'Mercy' does seem a term more appropriate to weighty matters, matters of life and death. But, though the *term* 'mercy' is not in great use, the *quality* does seem to me worth distinguishing and setting forth in its own right.

III

I suggested at the outset that there was a parallelism between both correct and incorrect accounts of mercy and forgiveness. I will now attempt to substantiate this with respect to the incorrect accounts.

Mrs. Smart's account of mercy and Dr. Kolnai's account of forgiveness both make the quality with which they deal appear paradoxical. On Mrs. Smart's account mercy at one end collapses into justice; at the other, is in danger of becoming a piece of injustice. Dr. Kolnai, in the section of his paper called 'The Paradoxy of Forgiveness' (pp. 95–9), sees forgiveness as at one extreme threatening to collapse into *condonation*, and at the other becoming *pointless*, for the wrongdoer has by his change of heart and reparation suitably annulled and eliminated his offence' (pp. 98–9), so there is nothing *to* forgive. This 'paradoxy' is in both cases (I suggest) a sign that something has gone amiss in the accounts.

What vitiates these accounts is their yearning to make mercy and forgiveness somehow *earned* or obligatory. They both insist that there must be *good reason* to be merciful or to forgive; for both the onus is on the merciful or forgiving man to account for, to *justify* his action. As Dr. Kolnai writes, 'Forgiveness is objectionable and ungenuine inasmuch as there is *no reason to forgive*, the offender having undergone no *metánoia* ("Change of Heart") but persisting in his plain identity *quâ* offender', p. 97 (emphasis in original). Dr. Kolnai's forgiving man, no less than Mrs. Smart's merciful one, faces his fellow man in the guise of (criminal) judge.

I will now attempt to sketch a satisfactory account of forgiveness.

IV

Holborow[1] writes 'The respect in which forgiveness represents a refusal to *hold* the faulty action *against* the player brings into clearer focus what is involved in blaming him . . .' Here, I believe, Holborow mistakenly takes forgiveness to be the opposite of blame. In doing so he is confusing blame and *resentment*. Resentment, like blame, does involve holding an action against someone, but one can only resent what one believes to be an offence done to oneself or persons near to one (family, close friends—people with whom one *identifies*, as we say). I can blame

[1] 'Praise, Blame and Credit', *P.A.S.*, 1961–2.

Nero for fiddling while Rome burned, but no one now living could resent Nero's action.

Forgiveness is opposed to resentment, not blame (cf. Butler, Sermon IX, section 2). To put things in a somewhat Kenny-an manner, one can only forgive what one could resent. It is this that most clearly distinguishes mercy and forgiveness; forgiveness is *essentially* a response to one who has offended you. (Thus I think Dr. Kolnai is confused when he says 'I should see a meaning in, say, forgiving the misdeeds of Napoleon I, but not those of Napoleon III . . . ' (p. 104)). It is important to separate blame and resentment, for one can forgive a person yet still blame him for what he has done. If this is so, it follows that there is not as great a danger of forgiveness collapsing into condonation as Dr. Kolnai suggests. The person you forgive may, perhaps rightly and inevitably, fall in your *esteem*, but (in another sense) you do not hold the offence against him, you "bear him no grudge", his action is no longer such an impediment to your relationship. As Dr. Kolnai observes, in forgiving him you *re-accept* him.

The parallelism with mercy I see in this way: by offending you a man, as it were, incurs a debt (hence we talk of owing recompense, reparation and apology). You are within your rights to resent his action. In forgiving him, you relinquish that right, you readjust your relationship to one of equality. This central feature is, I believe, distorted by Dr. Kolnai's judicial model.

The right to resentment may appear a highly suspect "right". And perhaps the use of this term does involve an extension of the notion. But not a totally alien one. We quite naturally stigmatize the presumption of one who offers to forgive a person who has not offended him by declaring that he has *no right* to forgive. We take strong exception to such behaviour because the man is assuming an unjustified superiority, an unfounded position of power. (It is perhaps in the light of such cases that we should consider the claim that 'forgiveness is the one unpardonable crime'.)

If this account is correct, in broad outline at least, we see the mistaken emphasis of Dr. Kolnai's account. On his view metánoia is seen as the *grounds* for forgiveness, it is as though a change of heart *earns* forgiveness. It is true that he does note the gift-like quality of forgiveness, but the burden of his account tells against it; he seems to incline to the view that forgiveness is in the central cases a (quasi-)duty: 'The fact remains that credible and perhaps "proven" change of heart constitutes the standard occasion to exercise and show forgiveness; it may be argued that genuine change of heart, and it alone, tends to make forgiveness a "duty" ' (p. 101). (Substitute 'stop blaming' for 'show forgiveness', and this may well be correct.)

But is there not some link between forgiveness and change of heart? There is, but not as the grounds for forgiveness, forgiveness is not

something to be *earned*. The link is far more oblique than Dr. Kolnai suggests. (i) Metánoia is a highly relevant factor in deciding whether to *tender* forgiveness: if the offender does not think he has offended to tender forgiveness will be to insult him, if he has not repented the wrong it may only harden him. (ii) Metánoia is a necessary condition for a sincere *request* for forgiveness: it is mere hypocrisy to ask for forgiveness while really unrepentant.

Forgiveness then is *gratuitous*, but that does not (of course) make it merely arbitrary. One's reasons for forgiving someone may be far more diverse, however, than simply the recognition of repentance. One may forgive an offence for old time's sake, or in the hope of a future unity, or because of a deep awareness of one's own weakness and liability to offend.

V

In conclusion, I would diagnose the view of morality which distorts the qualities of mercy and forgiveness as an obsession with duty and obligation. These are essential moral notions, but they are not the only ones. In their proper sphere they have valuable force and precision. But there is a tendency to allow them to monopolize (a tendency enshrined in the classification of generous, merciful, forgiving and kind acts as 'duties of imperfect obligation'). It is, I would claim, an impoverishment of moral life always to harp on about duty and obligation. Many highly moral actions are not *demanded* of one; they are not *earned* by their recipients, nor are they acts to which one is *bound*. Rather they are *gifts*, actions freely performed, sacrifices freely made. Eminent among these are the gifts of mercy and forgiveness.

University of St. Andrews

IMPLICIT RACISM

By SARA ANN KETCHUM *and* CHRISTINE PIERCE

ALTHOUGH there are no papers in recent philosophical literature which explicitly defend racist policies, several articles which defend sexist practices are implicitly racist. There are differences between the problems of racism and sexism—centring around reproductive function and sexuality—but these articles defend sexism on the basis of principles that apply to the areas where sexism and racism are similar.

At least three common defences of sexist practices are implicitly racist:

The Conservative Defence

According to the conservative, established patterns of prejudice should be respected by law. J. R. Lucas[1] claims that the following examples of sexist practices are morally justified and should not be corrected by law or policy:

> (1) Although it might be the case that working men as readily take orders from a fore-woman as a fore-man, or that customers would be as pleased to find a handsome boy receptionist as a pretty girl, there is no reason to suppose that it must be so . . . we recognise that a person's sex can reasonably be regarded as relevant to his or her suitability for particular posts, and that many institutions will operate on this basis, and are entitled to. I am justified in refusing to employ . . . a female foreman (pp. 166, 168).
>
> (2) . . . to deprive people of their liberty on any grounds irrelevant to their own desert is wrong: but it is not so evidently wrong to frustrate Miss Amazon's hopes of a military career in the Grenadier Guards on the grounds not that she would make a bad soldier but that she would be a disturbing influence in the mess room... Even if Miss Amazon is sure not to attract sidelong glances from the licentious soldiery, her sisters may not be; and it may be easier to operate an absolute bar than leave it to the recruiting officer to decide whether a particular woman is sufficiently unattractive to be safe (p. 167).

Lucas purports to use these examples in order to discuss when sexual differences justify institutional discrimination. However, he assimilates the sexist attitudes on the part of the workers to a lack of skill or ability in the fore-woman; similarly in the Miss Amazon case he treats the woman's attractiveness as a disqualification rather than treating self-control in the presence of attractive women as a qualification to be

[1] J. R. Lucas, 'Because You Are a Woman', *Philosophy*, 48 (1973), pp. 161–171. We are using Lucas's article to illustrate two common types of sexist defence; for a discussion of the particulars of his article, see: Susan Haack, 'On the Moral Relevance of Sex', *Philosophy*, 49 (1974), pp. 90–95; Trudy R. Govier, 'Woman's Place', *Philosophy*, 49 (1974), pp. 302–309.

required of the men involved. Thus, he in fact presents a defence of these cases on the basis of the prevalence of attitudes connecting the given role with a particular sex. He implies that if attitudes and habits of discrimination are sufficiently integral to the culture, discriminatory policies and laws are justified.

> The state is justified in not countenancing homosexual "marriages" because of our general understanding of what marriage really is, and the importance we attach to family life. For exactly the same reasons, women are debarred from being regarded in a fatherly or husbandly light; and hence also in those parts of the Christian Church that regard priests as being essentially fathers in God from being clergymen or bishops (p. 168).

If a widespread association between the role of priest or the role of soldier and maleness justifies excluding women from professions related to those roles, similar associations would justify excluding Blacks. There are many Whites who would object as much to a Black foreman as to a woman foreman on the grounds that they think of boss-servant roles in racial terms. A racist who thinks of servility and deference as essential characteristics of negritude will have as much difficulty taking orders from a Black boss as will a sexist who thinks of a woman boss as inappropriate on grounds that these same characteristics are essential to femininity. On the conservative argument policies directed against racism or sexism can be justified only if the society is not racist or sexist.

The Pseudo-Liberal Defence

According to the pseudo-liberal, equal opportunity for the sexes must wait on further investigation into sex differences. If certain treatment would be justified on the basis of characteristic X, and it is discovered that a considerably higher proportion of women than men are X, it would, on this argument, be justified to adopt a policy according such treatment to all women and only women. W. T. Blackstone[1] claims that Shulamith Firestone's rejection of differential treatment according to gender 'would be correct only if all relevant facts, characteristics, or circumstances which could in principle justify differential treatment were independent of gender' (p. 247). By 'independent of gender' he seems to mean that there is no correlation (or, at least, no sizable correlation) between that characteristic and gender. Lucas makes a similar claim: 'The more integrally and the more invariably a difference is connected with a person's sex, the more we are entitled to insist that the mere fact of being male or female can constitute a conclusive reason against being allowed to do something' (p. 167).

Certainly statistical differences between the sexes would indicate that

[1] William T. Blackstone, 'Freedom and Women', *Ethics*, vol. 85, no. 3, April 1975, pp. 243–248.

justice does not require equal distribution of women and men or Blacks and Whites within given job categories. But both Blackstone and Lucas are arguing against formal equality or equality of opportunity. According to their principles, all women may justifiably be discriminated against in hiring if sufficiently fewer women than men are qualified. Lucas argues for institutionalizing sexist discrimination in such cases on the analogy of the speed limit:

> We lay down rigid speed limits because they are easier to apply. There are many towns in which to drive at 30 mph would be dangerous, and many suburbs in which to drive at 45mph would sometimes be safe. But the advantage of having a fixed speed limit... outweighs its admitted unfairness (p. 167).

He compares the "exceptional" woman who actually is qualified to perform a masculine job to the highly-skilled driver who can drive safely at high speeds. Just as it is justified on the grounds of public good to prohibit the super-driver from driving fast, it may be justified on the same grounds to prohibit the super-woman from entering a masculine profession. Here he is assuming that the damage done to the women who are excluded from their chosen career and to the society which loses their skills is as trivial as the damage done to the good driver who is forbidden to drive at 100 m.p.h.

Just as Lucas's and Blackstone's principles would support a policy of refusing a woman a job solely on the basis of her sex where a sufficiently high percentage of that sex lack the qualifications, they would justify discriminating against Blacks in similar cases. Suppose that job X requires the ability to lift 30 lbs. and job Y requires an IQ of 140. Suppose further that 80% more men than women can lift 30 lbs. and 80% more Whites than Blacks have IQs of 140 or over. Then, if 80% is considered to be a sufficiently high correlation, it may be justified, according to the pseudo-liberal defence, to adopt a policy which would exclude all women from job X and all Blacks from job Y. Such a policy would require an employer to choose a male 98 lb. weakling over a strong woman for job X and a White with an IQ of 100 over a Black with an IQ of 150 for job Y. Neither Blackstone nor Lucas explains why he thinks that hiring people on the basis of sex—and, within the favoured sex, on qualifications—would be so much easier and more beneficial than hiring on the basis of individual qualifications.

The Libertarian Defence

Like the conservative, the libertarian argues against disturbing already existing prejudices by political means. W.E. Cooper[1] in his reply

[1] W. E. Cooper, 'What is Sexual Equality and Why Does Tey Want It?', *Ethics*, vol.85, no. 3, April 1975, pp. 256–257.

to Alison Jaggar's paper on sexual equality[1] says:

> My point is that the inequalities Jaggar deplores may arise because of the way that free men and women choose to lead their lives. *If* their choices reflect their conviction that the difference between the sexes is more than simply a physiological distinction, and indeed that the difference makes it fitting for them to introduce inequalities in advantages, *then* they should not be forced to conform to a feminist egalitarian vision (p. 256).

Unlike the conservative defence, the libertarian defence is based on freedom of choice; it is the fact that the institutions are freely chosen, rather than that they are integral to parts of the culture, which justifies them. Cooper cites Nozick's theory to illustrate this point: 'in criticizing socialist doctrine Nozick has argued that "the socialist society would have to forbid capitalist acts between consenting adults". In a similar vein I am arguing that Jaggar's feminist society would have to forbid sexist acts between consenting adults' (p. 257).

Cooper seems to be relying on the claim that no *moral* criticism can be made of cultural patterns: 'Perhaps [sexists] can be taught to change their minds. But the foundation for persuasion ought not to be the claim that they are acting unjustly, but rather, for example, that they would be happier in a sexually egalitarian society' (pp. 256–7). He implies that, as long as the members of the disadvantaged sex are convinced, by whatever means, of their inferiority or are unwilling to go to the trouble of asserting their equality against cultural odds, their oppression is just. Thus, one could produce a racially just society by convincing Blacks that they are better off as slaves, servants and manual labourers than they would be if they had opportunity equal to those of Whites to become doctors, politicians, etc. Since they would then consent to their lower status, this would be a racially just society.

Cooper does not state his examples as applying to any particular sex. However, he is defending the rights of those disadvantaged by sexism, and the examples he gives are unlikely, in the present society, to be problems for men. Hereafter, any specification of sex in our discussion of Cooper should be taken in the context of patriarchal sexism and be reversed for matriarchal sexism.

Cooper seems to think that, in order to defend the rights of women to choose a lowly position, he must defend the right to engage in sexist practices and that such practices can be set up and maintained without disadvantaging those women who do not want to be disadvantaged. He claims that Jaggar's principle—'that those of one sex, in virtue of ter sex, should not be in a socially advantageous position vis-à-vis those of the other sex' (p. 256)—may have the following problem:

> ... A is in a socially advantageous position relative to B if A has greater income than B. Now suppose that this situation obtains: B does not

[1] Alison Jaggar, 'On Sexual Equality', *Ethics*, vol. 84, no. 4, July 1974, pp. 275–291.

> want the higher income level of A, and B does not want it precisely because of ter sex. According to Jaggar's view there is an injustice here, and the only way to rectify it would presumably be to force B to accept the higher income against ter will. But surely this coercion is the real injustice, violating B's right to lead ter life as tey choose (p. 256).

Jaggar's principle does not say anything about what means to achieving the non-sexist society would be justified, so Cooper's complaint about coercion is simply off the point. Moreover, if the woman in this example is unambitious (for whatever reason), that may be reasonable non-sexist grounds for her to be at a disadvantage. In that case, she would be at a disadvantage as a member of the class of people who are not ambitious rather than as a woman. There is no conflict between the policy of discriminating against unambitious people and Jaggar's principle, and no need to establish sexist practices in order to protect the rights of unambitious women or men.

In defending people's freedom to set up and participate in unequal institutions, Cooper overlocks the fact that such institutions regulate and affect people other than those who choose them. For example, he defends the right of a woman to be discriminated against as a woman if she wants to because, for instance, she is in love with a sexist spouse. However, one person cannot be discriminated against as a woman without there being some practice or policy of discrimination which would affect other women who might not so choose. The same principle would defend the right of a Black to his servility and lower status if he wants to preserve a good relationship with the racist Whites in his town. Again, assuming that he is not the only Black around, his lower status as a Black is logically tied to the lower status of other Blacks who may not consent to it. No articles in the literature champion the rights of Blacks to be treated as inferior (or to be slaves if they want to be) in order to safeguard their freedom, and Cooper offers no reason for granting such rights to sexists while not granting them to racists.

All three articles claim to be defending sexist practices for the benefit of women; hence rescuing them from the follies of their more political sisters. Cooper sees nothing problematic in advocating the freedom to be sexist as a defence of women's rights. Nor does Lucas see anything strange in proposing the institutionalization of sexist prejudice; and both Lucas and Blackstone imply that women ought to be protected from the freedom to undertake the hazards of doing the same kind of work that men do. Perhaps a higher sensitivity makes us more suspect of arguments which defend racism on the grounds that it is good for Blacks.

State University of New York, Oswego

ENGLISH AND TRUTH FUNCTIONS

By E. J. Borowski

KEITH HALBASCH in his recent paper (Analysis 35.3) remarks that ⌜Neither φ_1 nor φ_2 . . . nor φ_n⌝ is true (for any n) just in case none of the constituents is true, and concludes that, since there is no binary bracketing for $n = 3$ which gives a sentence equivalent to the given one, this is an example of a ternary sentence connective in English. He then generalizes the argument for any n, and so concludes that there must be countably many 'truth functional words (or morphemes) . . . in English'. Unfortunately it is clear that he has followed a red herring to this conclusion while overlooking a simpler but more serious deficiency.

An apparently trivial negative point is that it is necessary to restrict discussion to $n \geqslant 2$, since ⌜Neither φ_1⌝ is not well formed in English. This parallels the inadmissibility of ⌜φ_1 nor φ_2⌝ of which Halbasch is aware, since he remarks that in order to obtain ⌜Neither φ_1 nor φ_2 nor φ_3⌝ by binary bracketing, we should have to assume the deletion of one 'neither'.

However, Halbasch claims to present a problem for empirical linguistics, yet his problem only is a problem if it is stated in terms of a naive grammar without consideration of possible transformational derivations of sentences of the form he considers. True, he mentions this possibility at the end of his paper, but it is not merely his priorities at which I cavil; it is rather the conclusion that the 'truth function structure of English is more complicated than normally supposed'. For what is 'normally supposed' nowadays is that surface English is derived from some deeper language, be it of language-independent semantic primitives or just quasi-English, in which the problem does not arise. It should be noted that I am not here concerned to discuss the possibility of there being a single semantically primitive truth function, such as the *N*-operator of the *Tractatus*, from which all others can be recursively derived, but only to show that at some relatively deep level each surface manifestation can be so represented that the number of arguments is immaterial while there remains a uniform syntactic derivation of the surface form from the deep one.

Certainly recursive rules are required at this level, but why should Halbasch imagine that there must be 'infinitely many precise transformations' recursively specified? Surely a small number of transformations applied recursively would be sufficient. Of course such a transformation could not simply take ⌜Not (A v B)⌝ into ⌜Neither A nor B⌝, since when A or B is itself of the form ⌜C v D⌝ we could not proceed by reapplying the same rule. However, the desired result is achieved simply by successive application of two transformations. First we make

a purely notational change, writing ⌜It is not the case that (φ_1 or φ_2 or . . . or φ_n)⌝ as ⌜NOT: OR (φ_1, φ_2, . . . , φ_n)⌝ where 'OR' may for the moment be regarded as an operator on a set of sentences, although this view will later be modified. (Note that lower case conjunctions are English words, while capitals represent abstract connectives.) This is now first transformed into ⌜Neither φ_1 NOR (φ_2, . . . , φ_n)⌝, where 'NOR' similarly takes a set as argument but is merely an auxiliary symbol in the syntactic transformation which will not appear in the output. This requirement is met by the recursive application of a rule transforming ⌜NOR (φ_m, . . . , φ_n)⌝ into ⌜nor φ_m NOR (φ_{m+1}, . . . , φ_n)⌝ and ⌜NOR (φ_n)⌝ into ⌜nor φ_n⌝. Clearly this is not a problem for empirical linguistics, nor does it involve the postulation of 'infinitely many precise transformations'. Furthermore, it allows for the derivation of the alternative English form ⌜Neither φ_1, φ_2, . . . , nor φ_n⌝ directly without recursion, a possibility which exists for 'or' and 'and' too, and which Halbasch ignores.

Another objection to Halbasch's conclusion is that it is not clear why we should expect English to have an overt truth-functional structure at all. At best we might expect that the semantic properties of sentence connectives of English should be explicable in terms of truth functions. (How else will we be able to derive the semantic properties of compound sentences from those of simple ones?) But there is no reason whatever to suppose that this discovered structure, which after all we impose on English in conformity with our logic, should be isomorphic to the surface grammar of English. Halbasch's thesis is not materially different from the claim that the occurrence of 'and' in English shows its truth-functional structure to be more complex than that of a logic with negation and implication as primitive connectives. It is clear that this has put the cart before the horse.

Such criticisms aside, I think it clear that Halbasch's conclusion can be strengthened: there are ordinary English connectives for which no adequate truth-functional account can be given. For consider ⌜Either φ_1 or . . . or φ_n⌝, which he says 'can be construed as a series of binary disjunctions'. Unfortunately, English does not always use disjunction inclusively in this way. Often instead a proposition of this disjunctive form (for $n \geqslant 2$) will be true if and only if *one and only one* φ_i ($1 \leqslant i \leqslant n$) is true. For example, consider issuing an invitation to 'the chairman, treasurer, . . . , or secretary' of some committee to attend a meeting: one would, I think, be justifiably aggrieved if more than one office bearer arrived, no matter how many posts were specified. Note that there is no intrinsic contextual reason for this disjunction to be exclusive, nor is it anything to do with the imperative mood. The minute of the previous meeting might read 'the chairman, treasurer, . . . , or secretary has been invited to attend' (—in which, we ought, *en passant*, to notice that a singular verb is required).

Now this time it is clear that there can be no uniformly iterable truth function which will be equivalent to this disjunction for all n. The problem is now a different and important one: not how to account for the connective for each n, but how to do so for all n. Clearly, we must suggest something which will yield the standard binary nonequivalence $\ulcorner \varphi_1 \underline{\vee} \varphi_2 \urcorner$ when $n = 2$. But this cannot itself just be iterated, since when $n = 3$, $\ulcorner (\varphi_1 \underline{\vee} \varphi_2) \underline{\vee} \varphi_3 \urcorner$ is true when φ_1, φ_2, φ_3 are all true. Nor can we use a similar transformation to that suggested above, taking $\ulcorner$NOT $(\varphi_1$ iff . . . iff $\varphi_n)\urcorner$ into $\ulcorner \varphi_1$ or . . . or $\varphi_n \urcorner$, since with $n = 3$, whether we bracket equivalences from the left or right, $\ulcorner \varphi_1$ or φ_2 or $\varphi_3 \urcorner$ will be false when φ_1 and φ_2 are true and φ_3 false, while $\ulcorner \sim((\varphi_1 \equiv \varphi_2) \equiv \varphi_3)\urcorner$ will be true. Here $\ulcorner$NOT: IFF $(\varphi_1, \ldots, \varphi_n)\urcorner$ will not fit the bill either, where $\ulcorner$IFF $(\varphi_1, \ldots, \varphi_n)\urcorner$ is true just in case all φ_i have the same truth value, since when $n = 3$, the negated sentence is true when one or two constituents are true. Similarly, while $\ulcorner \varphi_1$ or $\varphi_2 \urcorner$ is equivalent to $\ulcorner (\varphi_1 \vee \varphi_2) \,\&\, \sim(\varphi_1 \,\&\, \varphi_2)\urcorner$ this cannot be straightforwardly generalized, since for $n \geqslant 3$, $\ulcorner (\varphi_1 \vee \varphi_2 \vee \ldots \vee \varphi_n) \,\&\, \sim(\varphi_1 \,\&\, \varphi_2 \,\&\, \ldots \,\&\, \varphi_n)\urcorner$ does not exclude the cases where more than 1 but less than $n - 1$ of the φ_i are true.

Of course $\ulcorner \varphi_1$ or φ_2 or . . . or $\varphi_n \urcorner$ is equivalent to $\ulcorner$only $\varphi_1 \vee$ only $\varphi_2 \vee \ldots \vee$ only $\varphi_n \urcorner$ since now the inadmissable conjunctions are ruled out by the form of the disjuncts, and a recursive definition can be given for the disjunction (for each n) as $\ulcorner (\varphi_1 \,\&\, \sim\varphi_2 \,\&\, \sim\varphi_3 \,\&\, \ldots \,\&\, \sim\varphi_n) \vee (\sim\varphi_1 \,\&\, \varphi_2 \,\&\, \sim\varphi_3 \,\&\, \ldots \,\&\, \sim\varphi_n) \vee \ldots \vee (\sim\varphi_1 \,\&\, \sim\varphi_2 \,\&\, \ldots \,\&\, \sim\varphi_{n-1} \,\&\, \varphi_n)\urcorner$. However, it is exceptionally implausible heuristically that a form as complex as this should underlie our use of a connective which we have no difficulty in understanding. In addition, this treatment makes no allowance for the addition of another disjunct as an afterthought, since n has to be fixed in advance of the analysis of even the first disjunct.

Furthermore there are compelling linguistic reasons for treating even 'AND' as an operator on sets of sentences rather than on only two at a time. Thus McCawley ('A Program for Logic', *Semantics of Natural Language*, ed. Davidson & Harman) argues that sentences of the form

(a) John φ'd and John ψ'd and Bill ψ'd

(b) John φ'd and he & Bill ψ'd

(c) John φ'd and ψ'd and Bill ψ'd

reflect the difference between iterated binary conjunction (*b* and *c*) and set conjunction. Also, intuitively, we feel justified in *un*bracketing iterated conjunction or inclusive disjunction just because the bracketing makes no difference, whereas iterated binary exclusive disjunction is a logical foundling, since 'there is no language with a conjunction which fits (its) truth table—i.e. is true when an odd number of its conjuncts are

true, but false when an even number of them are true'. But the most startling point of all is that conjunction and inclusive disjunction now just are universal and existential quantification over finite sets of sentences, which, of course, as McCawley points out, considerably simplifies the statement of logical rules, since, e.g., the identification of conjunction and universal quantification allows the statement of a single rule, licensing the derivation of φ_i ($1 \leqslant i \leqslant n$) from $\ulcorner$AND ($\varphi_1, \ldots, \varphi_n$)$\urcorner$, in place of both conjunction and universal elimination. And this too is not merely an abstract result, but one for which McCawley finds syntactic evidence.

This might seem to support the remarkable conclusion that the empirical facts of language with regard to multiply connected sentences actually support the *Tractatus* view of the relation between the quantifiers and the propositional connectives. Thus, for example, there might be alternative realizations of 'AND', 'OR' (inclusive), and even 'NOR' as 'all of', 'at least one of' and 'not/nor any of' respectively. But even laying aside the difficulty that all of these must be followed by 'AND' again and that this 'AND' must be realized in the reduced form, since polysyndeton would here be ungrammatical for the reasons already given, such an account of the exclusive 'OR' (henceforward 'AUT') is not possible while preserving the equivalence of the successive recursive products. The problem here is not a purely syntactic one but one of the relation of syntax to semantics. However, this case displays an even deeper problem, which is sufficient to show that, even if some abstract operators, such as Wittgenstein's '*N*', can be so represented, not all can.

The arguments of the sentential connectives cannot be sets, since, although the set $\{\varphi, \varphi, \sim\varphi, \sim\varphi\}$ is just the set $\{\varphi, \sim\varphi\}$, the exclusive disjunction $\ulcorner$AUT ($\varphi, \varphi, \sim\varphi, \sim\varphi$)$\urcorner$ is necessarily false while $\ulcorner$AUT ($\varphi, \sim\varphi$)$\urcorner$ is necessarily true. Formally, to deal with this we distinguish the two occurrences of φ but stipulate that they have the same truth value, but this is intuitively unsatisfactory. (Which does $\ulcorner\sim\varphi\urcorner$ negate?) A new notion intermediate between Boolean sets and ordered sets could be defined simply *ad hoc*; in these "occurrence sets" the order of elements would be immaterial, but sets with different numbers of occurrences of the same element would be distinct. However, there is an immediate problem in the definition of union and intersection for these, namely that we would have no way—without resorting to the Boolean subscripts anyway—of deciding whether occurrences of an element in different sets were the same or different. So, although such a notion is heuristically useful in representing the prime factors of integers whose highest common factor or least common multiple is to be calculated, or numbers whose average is sought, the formal difficulties outweigh the intuitive simplicity. In any event, in the present context it is not even clear that the introduction of the new concept is well motivated; for if the only problem is with sentences of the form $\ulcorner$AUT (φ, φ, ψ)$\urcorner$ the curious fact

MEMORY UNCHAINED AGAIN

By JOHN A. SCHUMACHER

SYDNEY Shoemaker ('Persons and Their Pasts', *American Philosophical Quarterly*, VII (1970), p. 282) claims that 'unless we understand the notion of retention, as well as that of memory, as involving a causal component, we cannot account for the role played by the notion of memory . . . in judgments of personal identity'. In this paper I shall argue that Shoemaker is mistaken, and more generally that any philosopher who takes the notion of memory to play some sort of special role that requires it to be given a causal analysis is equally mistaken. I begin by reviewing some earlier contributions to our understanding of the notion of memory.

According to B. S. Benjamin ('Remembering', *Mind*, LXV (1956), pp. 321 and 329), a demonstration that a person—say, A—has not forgotten an event—say, X—is that he has produced an accurate representation of X. Benjamin's use of the term 'forgotten' invites us to make a connection between A's present representation of X and his past experience of it. But can A's presently exercising an ability to represent X tell us enough about how he acquired that ability? Can it tell us, for example, that he acquired it by experiencing X? Ever since C. B. Martin and Max Deutscher's response to Benjamin ('Remembering', *Philosophical Review*, LXXV (1966), Section V), the answer to these questions has obviously been 'no', for Benjamin simply fails to consider, and thus to account for, what happened to A between his past experience of X and his present representation of it.

One of Martin and Deutscher's examples concerns a painter who believes he is painting a completely imaginary scene which turns out to be an accurate reproduction of a scene he once viewed in his childhood. According to Martin and Deutscher, we should resist saying that this is a case of memory until we can conclude that the 'only reasonable explanation' for his painting the scene is that he remembers it from his childhood. However,

> if to remember an event . . . were merely to represent it and to have observed it, then it would be absurd to pretend to *explain* the fact that someone gave a description which fitted something he had seen, by asserting that he must be remembering it (p. 176).

But what sort of connection is now required between his past observation of the event and his present representation of it? Martin and Deutscher answer

> Whether he has been told about it in the meanwhile, how young he was at the time, whether he had seen things very similar at many other times

> are all relevant to deciding whether he actually remembers the event. These facts are the same as those which are used to decide whether or not he would have given the story if he had *not* witnessed the event in his childhood. To decide that he would not have done so is to decide that his past witnessing is causally necessary for his present account (p. 176).

For Martin and Deutscher, then, the connection is causal: A's past observation of X must be operative in producing a state or successive states in him finally operative in producing his representation. (The term 'operative' is introduced in order to allow a causal condition to be neither necessary nor sufficient (p. 179), but I shall not discuss this here.)

On the other hand, Roger Squires ('Memory Unchained', *Philosophical Review*, LXXVIII (1969)) argues that all we need to say to connect A's present representation of X to his past observation of it is that, in virtue of having observed X, A acquired an ability to represent X, an ability which he retained and now exercises in representing X. Squires asks, 'Could we not say that, although he [the painter] did not realize it, he acquired and kept a capacity to paint certain scenes from his childhood' (p. 191)? Clearly, for Martin and Deutscher, saying this would not be enough; retaining such a capacity would have to be analysed in causal terms. But Squires counters this by arguing that 'in describing something as having retained a quality, as having stayed the same in a certain respect [e.g., by retaining some ability], there is no forced reference to causal connections' (p. 178); there need be no causal connection between its previous state(s) and its present one.

> We can sometimes explain the fact that someone who had a car last year has a car now by saying that he must have had it all along. But this is not a causal explanation. Similarly, when we explain that someone must be remembering we are, in general, either pointing out that what he did was a display of knowledge (rather than an accident, say) or we are saying that he has had that knowledge all along (p. 191).

The conditional 'If A had not had his ability to represent X yesterday, he would not have had it today' is not a causal conditional; in asserting it we would be denying that certain causal explanations (such as, perhaps, that A had observed X this morning) are necessary, by eliminating the possibility that A had acquired his ability since yesterday. Similarly, with respect to the conditional 'If A had not observed X in his childhood, he would not have represented X as he did today', we would be eliminating the possibility that A had acquired his present ability to represent X since he observed X in his childhood. (I shall eventually say more about acquiring and exercising abilities.)

Nevertheless, Sydney Shoemaker does raise some doubts about Squires' argument; for the reason I indicate above, he claims that we must follow Martin and Deutscher in giving a causal analysis of the

retention of an ability to represent an event. Shoemaker bases his claim on the following example:

> Let us suppose that the brain from the body of one man, Brown, is transplanted into the body of another man, Robinson, and that the resulting creature—call him 'Brownson'—survives and upon regaining consciousness begins making memory claims corresponding to the past history of Brown rather than that of Robinson. We can also suppose that Brownson manifests personality traits strikingly like those previously manifested by Brown and quite unlike those manifested by Robinson. . . . I think that most people would want to say that Brownson is (is the same person as) Brown (p. 282).

Shoemaker argues, in effect, that the notion of the retention of abilities cannot do the required work here, for the assertion that Brownson retains abilities acquired by Brown in the past presupposes the identity of Brown and Brownson and, consequently, 'cannot without circularity be offered as evidence for it' (p. 282). Rather than using this notion, Shoemaker advocates asserting that Brownson's present abilities are *causally related* to Brown's past actions and experiences in essentially the same ways as a normal person's present abilities are causally related to his own past actions and experiences. I should like to argue, on the other hand, that asserting this is not at all necessary, and that Shoemaker has not seen the actual force of both Squires' and his own arguments.

First, following Shoemaker's definition of quasi-remembering (p. 271), let me define the quasi-retention of abilities of some person, say, A: A's quasi-retaining an ability requires that there be a correspondence between his present state in which he has that ability and *a* past state in which that ability was acquired, but the past state need *not* be a past state of A himself, although otherwise the correspondence is *just like* that which exists in the normal retention of abilities. With respect to quasi-remembering, Shoemaker jumps from saying that the correspondence in question must involve only a matching of content between the past and present states to saying that it must involve a causal connection. He shows that the former is inadequate in important ways, but, clearly, there can be intermediate positions, one of which involves the usual facts about the retention of abilities. Martin and Deutscher themselves provide examples of such facts in a passage I quote above.

Secondly, I do not beg the question of identity by asserting that Brownson quasi-retains abilities acquired by Brown in the past, and thus this assertion can be used as evidence for the identity of Brown and Brownson, just as the assertion that Brownson quasi-remembers Brown's past actions and experiences can be used. And, thirdly, I can use Squires' argument to show that, in describing the correspondence involved in quasi-retention, there is no *forced* reference to causal connections: the conditional 'If Brown had not acquired certain abilities,

Brownson would not have had those abilities' is not a causal conditional, for to assert it is to deny that certain causal explanations (such as, perhaps, that Brownson acquired those abilities *after* the operation) are necessary, by eliminating the possibility that those abilities were not acquired by Brown. This sounds odd because in our world quasi-retention is always retention, but if Shoemaker can say what he does about quasi-remembering and remembering, I see no reason why I cannot say what I do here.

One more potentially puzzling feature of this resolution still remains, though. No doubt we can feel some residual concern about the obviously causal nature of transplanting brains. But why do we allow the notion of retention to bear the brunt of this concern? Looking back to the case of remembering, once we remove the tendency to see the causal component as entering via the notion of the retention of an ability, it is open to us to see it as entering via the notion of acquiring, exercising, or merely having an ability. I take it that this is the central project of, for example, D. M. Armstrong in *A Materialist Theory of the Mind* (Chapter 6, Sections 5 and 6): 'the concept of a mental state is primarily the concept of *a state of a person apt for bringing about a certain sort of behaviour* [and secondarily, in some cases] *apt for being brought about by a certain sort of stimulus*'. Thus, for Armstrong, in describing someone as being disposed, or as being able, to do something there is *already* a forced reference to some sort of causal mechanism, in this case a state with causal powers. We should note, of course, that one way the person can stay the same is by retaining such a state, so that no *further* causal mechanisms or connections are required to account for retention here.

However, even though we might be able to salvage a causal analysis of remembering in this way, and thereby to resolve our concern about transplanting brains, we would also be carried far beyond a mere analysis of remembering and into a causal analysis of cognitive abilities in general. These two analyses work at different levels: we can analyse remembering simply in terms of acquiring, retaining and exercising an appropriate ability, and if it turns out that Armstrong is right in thinking that in describing someone as being able to do something there is a forced reference to some sort of causal mechanism, then we should incorporate this particular analysis of having abilities into our analysis of remembering, but this could in no way undermine what we had already accomplished in our analysis. Nor could it undermine my rejection of Shoemaker's doubts about unchaining memory.

Rensselaer Polytechnic Institute

ON A PRIORI CONTINGENT TRUTHS

By W. R. CARTER

I WILL try to show that there is a simple, but telling, objection to Saul Kripke's claim to have provided counter-examples to the thesis that *a priori* truths are invariably necessary truths. Although I will address only one of Kripke's examples, I believe the point to hold for them all.

Suppose that a *metre* is defined as the length of a certain stick, *S*. Kripke is careful to note that the definition is employed 'to fix the reference' of the word 'metre' and not to give its meaning ('Naming and Necessity', in *Semantics of Natural Language*, ed. by Donald Davidson and Gilbert Harman, D. Reidel, 1972, p. 274). Consider the statement that *S* is one metre long. Kripke claims (a) that this statement is, at least for the person who gives the definition of *metre* just cited, an *a priori* truth, and (b) that this statement is contingent and not necessary, since *S* might have been a different length than it is. If both claims can be sustained then it is clearly false that all *a priori* truths are necessary. I will argue against (a) but not (b).

Is the statement that *S* is one metre long an *a priori* statement? Kripke seems to me right to suggest that what is *a priori* for one kind of creature may not be *a priori* for another. To say that something is an *a priori* truth is presumably to say something like this: that it can be known independently of sense experience (p. 260). But what Martians, or God, can know independently of sense experience may be quite different from what people can know independently of sense experience. Moreover, Kripke suggests that what is *a priori* for someone who defines *metre* in the manner sketched earlier may not be *a priori* for the rest of us. I will argue, not that the statement that *S* is one metre long is not *a priori* for God or Martians, but that it is not *a priori* for human beings, including those human beings who define words.

The argument begins with the premiss that if it is true that an object has a property then there exists such an object. More generally, if an object has a property in the context of a possible world *W* then there exists such an object in *W*. Given this, it follows that one of the things that must be true if it is to be true that *S* has the property of being one metre long is that there exists such an object as *S*. Thus before we can know that it is true that *S* has the property of being one metre long, or has any other property, we must know that there is such an object as *S*. But our knowledge that there is such an object is clearly not *a priori*, is not the sort of knowledge that is independent of sense experience. More precisely, the statement that there is such an object as *S* is not an *a priori* truth for human beings, including those human beings who define metre in the Kripkian manner, even if it is *a priori* for God. Our definers first

observe object *S*, then go through some sort of ceremony whereby they define a word in terms of *S*. Once they have done all of this they do not need to check anything, or observe anything, to know that *S* is one metre long. But this does not justify saying that the statement that *S* is one metre long is, for our definers, an *a priori* truth. We might as well say that the statement that my fountain pen is on my desk is for me an *a priori* statement, this on the grounds that once I have observed the location of the pen I do not *then* have to rely upon sense experience to know that my pen is on my desk.

North Carolina State University at Raleigh

PLANTINGA AND THE CONTINGENTLY POSSIBLE

By HUGH S. CHANDLER

ONE of the interesting things in Alvin Plantinga's latest book, *The Nature of Necessity* (Oxford, Clarendon Press, 1974), is his contention that there are no states of affairs which are only *contingently* possible, i.e. possible in some possible worlds and not possible in others. This contention is important to Plantinga. His new version of the ontological argument is unsound if the contention is false. Nevertheless I want to suggest that what is possible varies from world to world. It should be noticed at the outset that the possibility in question is possibility 'in a broadly logical sense', to use Plantinga's words; not, for example, physical possibility.

I

We intuit that a particular bicycle which, in fact, came into existence made up of parts P_1, P_2, P_3, . . . P_n, could not have come into existence made up of totally different parts. On the other hand, this bicycle *could* have come into existence with *one* of its parts different from the one it actually had. For example, our bicycle could have been constructed with a spoke other than one of its actual front-wheel spokes. (The bicycle-builder says to his assistant, 'If you had ordered new spokes as I told you, this would have been a better bicycle'.)

Consider a possible world in which the bicycle now before us came into existence with a different spoke. Surely *that* bicycle could have come into existence with, say, a different handle-grip than one of the ones it did have at its origin, there, in that possible world. That is to say, we now see a second possible world in which the bicycle I indicated in the

first possible world came into existence with a different handle-grip. Proceeding in this fashion, we seem to work our way towards a world in which our bicycle came into existence made entirely of parts other than $P_1, P_2, P_3, \ldots P_n$. The bicycle with which we began is one and the same as the one in the first non-actual world, and that, presumably, is one and the same as the one in the second, and so on. But, given our intuition, how can the first and last bicycle in this series be the same? We seem forced to hold that identity across worlds is not transitive, i.e. is not *identity*.

David Lewis would say that all these bicycles are linked by *resemblance*.[1] The bicycle in the first non-actual world is the 'counterpart' there of our bicycle—that is, it is the bicycle which, in that world, most nearly resembles our bicycle. The counterpart relation is not transitive. Thus Lewis has an easy escape from the puzzle about the bicycle. Unfortunately Plantinga, among others, seems to have shown that the counterpart theory is unacceptable.[2]

I want to offer a different solution to the problem about the bicycle. But first, to simplify matters, consider an object made up of just three elements—i.e. an 'alpha'. Suppose that a certain alpha (call it 'Alfred') came into existence ten minutes ago made up of elements E_1, E_2, and E_3. And suppose that the rules governing the re-identification of alphas permit us to say that Alfred could have been made up originally of E_4, E_2, and E_3, or of E_1, E_5, and E_3, or of E_1, E_2, and E_6, but they ordain that it is essential that Alfred have been made up of at least two of the three elements of which it was, in fact, composed.

Since Alfred, in fact, came into existence made up of E_1, E_2, and E_3, it is necessary that it should have been made up of at least two of these elements. But, if it had come into existence (and this was a possibility) made up of E_4, E_2, and E_3, then it would have been possible for it to have come into existence made up of E_4, E_5, and E_3. That is to say, it could have lacked a property which, in fact, it necessarily has.

Should the advocates of possible worlds say that there is a possible world, w_2, in which Alfred came into existence made up of E_4, E_5, and E_3? Perhaps they should say that, relative to the actual world, w_0, there is no such world. But there is a world, w_1, such that, relative to *that* world, w_2 is a possible world. (Presumably, in w_1, Alfred came into existence made up of E_4, E_2, and E_3.) How else can our claims about what would have been possible be analysed into talk about possible worlds?

There is a difficulty.[3] Surely there is a possible world—possible with respect to w_0—in which Alfred never exists, but some other alpha

[1] David Lewis, 'Counterpart Theory and Quantified Modal Logic', *Journal of Philosophy*, March 7, 1968.

[2] *The Nature of Necessity*, pp. 102–120. Also Fred Feldman, 'Counterparts', *Journal of Philosophy*, July 1, 1971.

[3] The problem that follows was suggested to me by Robert Stalnaker.

(Bernard) comes into existence in the same place, made up of E_3, E_4, and E_5. Call this world 'w_3'. We assume that w_3 is exactly like w_2 except for the fact that in w_2 a certain alpha is Alfred, while in w_3 the corresponding alpha is Bernard. But now we need to reconsider what we have just said. Notice that in both w_2 and w_3 exactly the same elements are arranged in exactly the same way in exactly the same place at exactly the same time, and thus an Alpha is brought into existence. How can this procedure create *different* alphas? Perhaps w_2 and w_3 are really the same world (and we are calling one alpha by two different names). On the other hand, perhaps w_2 and w_3 are different worlds (and Alfred and Bernard different alphas) simply because those worlds stand in different relations to the actual world.

Let's get back to the bicycle. A bicycle can survive the gradual replacement of each and every one of its parts. But, given that the bicycle came into existence made up of P_1, P_2, P_3, . . . P_n, it is not possible that it should have come into existence made up of entirely different parts. It could have originated with one of these parts replaced, or with two perhaps, with three, but not with a total replacement. Where, in this series, do we pass (*via* a region of indeterminacy) from possibility to impossibility?

Pretend that our bicycle could have come into existence with half of its parts replaced, but could not have come into existence with two-thirds of its parts replaced. Under this hypothesis, 'it is not possible that this bicycle should have come into existence made up of entirely different parts' cannot be interpreted to mean that there are no worlds of any sort in which the bicycle came into existence made up of entirely different parts. It can only be taken to mean that worlds in which this occurs, if there are any, are not possible relative to the *actual* world. For, after all, our bicycle could have come into existence with half of its parts replaced. And, if that had been the case, it would have been possible that our bicycle should have come into existence made up of parts entirely different from P_1, P_2, P_3, . . . P_n.

Now assume that our bicycle could have originated with any third of its parts replaced, and could not have originated with half of its parts replaced. Under this assumption, we have to cross from a would-have-been-possible world to a would-have-been-a-would-have-been-possible world in order to find "our bicycle" with all of parts P_1, P_2, P_3, . . . P_n replaced. I suspect that this cannot be done.

Our bicycle could have come into existence with one third of its parts replaced. And, if it had originated in that condition, it would have been possible for it to have come into existence with two thirds of its parts other than P_1, P_2, P_3, . . . P_n. We have moved from a possible world to a would-have-been-possible world. Next one wants to say 'Suppose our bicycle had come into existence with two thirds of parts

P_1, P_2, P_3, . . . P_n replaced—in that case, a total replacement would have been possible'. But, by hypothesis, we *cannot* suppose that our bicycle came into existence with two-thirds of parts P_1, P_2, P_3, . . . P_n replaced. Apparently, we can take one step into the realm of the impossible, but not two.

II

Why does Plantinga believe that what is possible does not vary from world to world? He offers the following argument.

> (1) If a state of affairs S is possible, then it is necessarily possible; that is, possible with respect to every possible world.

From this, he claims, it follows that

> (2) Every possible world is possible with respect to every possible world.

And from this he derives

> (3) Any state of affairs possible with respect to at least one possible world, is possible with respect to every possible world. (Ibid., p.54.)

No evidence is offered for (1). One has the impression that it is supposed to be intuitively obvious. My intuition is that it is false.

Alfred came into existence made up of E_1, E_2, and E_3. But it could have come into existence made up of E_4, E_2, and E_3, or of E_1, E_5, and E_3. Thus the state of affairs of Alfred's having come into existence made up of E_4, E_2, and E_3 is possible; but it is not *necessarily* possible. There is a possible world in which Alfred came into existence made up of E_1, E_5, and E_3, and in *that* world it is *not* possible that Alfred should have come into existence made up of E_4, E_2, and E_3. The rules governing the re-identification of alphas forbid it. Hence it is only a *contingent* fact that this state of affairs is possible.

University of Illinois at Champaign-Urbana

THE CIRCULARITY OF THE PROOF OF THE NON-INDEPENDENCE OF THE FOURTH AXIOM OF *PRINCIPIA MATHEMATICA*

By JOSEPH DANQUAH

IN this note we establish that Bernays's classical proof that Axiom A4 (*1.5) of *Principia Mathematica*'s propositional calculus is redundant under substitution and detachment is circular and rests on a covert use of A4. The circularity would have been avoided if his proof had been preceded by a proof of *Syll.* without the aid of A4. No such proof is given in Bernays's paper (Paul Bernays, 'Axiomatische Untersuchungen des Aussagenkalküls der *Principia Mathematica*', *Mathematische Zeitschrift*, vol. 25, 1926, pp. 305–320). Bernays uses the following abbreviated names of the five axioms of PM's PC which we designate A1, A2, A3, A4 and A5.

A1	$(p \vee p) \supset p$	*Taut*
A2	$q \supset (p \vee q)$	*Add*
A3	$(p \vee q) \supset (q \vee p)$	*Perm*
A4	$(p \vee (q \vee r)) \supset (q \vee (p \vee r))$	*Asso*
A5	$(q \supset r) \supset ((p \vee q) \supset (p \vee r))$	*Sum*

Bernays's proof, which appears on pp. 312, sq., of his paper, is reproduced below to facilitate reference.

(*Add*)	$r \supset p \vee r$	[B1]
(*Sum*)	$r \supset p \vee r. \supset . q \vee r \supset q \vee . p \vee r$	[B2]
	$q \vee r \supset q \vee . p \vee r$	[B3]
(*Sum*)	$q \vee r \supset q \vee . p \vee r. \supset . p \vee . q \vee r$ $\supset p \vee : q \vee . p \vee r$	[B4]
	$p \vee . q \vee r \supset p \vee : q \vee . p \vee r$	[B5]
(*Perm*)	$p \vee : q \vee . p \vee r \supset q \vee . p \vee r : \vee p$	[B6]
	$p \vee . q \vee r \supset q \vee . p \vee r : \vee p$	[B7] (1)
(*Add*)	$p \supset r \vee p$	[B8]
(*Perm*)	$r \vee p \supset p \vee r$	[B9]
	$p \supset p \vee r$	[B10]
(*Add*)	$p \vee r \supset q \vee . p \vee r$	[B11]

	$p \supset q \vee . p \vee r$	[B12]
(*Sum*)	$p \supset q \vee . p \vee r . \supset . q \vee . p \vee r : \vee p \supset q$ $\vee . p \vee r : \vee : q \vee . p \vee r$	[B13]

	$q \vee . p \vee r : \vee p \supset q \vee . p \vee r : \vee : q \vee . p \vee r$	[B14]
(*Taut*)	$q \vee . p \vee r : \vee : q \vee . p \vee r \supset q \vee . p \vee r$	[B15]

	$q \vee . p \vee r : \vee p \supset q \vee . p \vee r$	[B16] (2)
(1)	$p \vee . q \vee r \supset q \vee . p \vee r : \vee p$	[B17]
(2)	$q \vee . p \vee r : \vee p \supset q \vee . p \vee r$	[B18]

	$p \vee . q \vee r \supset q \vee . p \vee r$	[B19]

Q.E.D.

The formulas appearing under the ruled lines are the derived lines by substitution and detachment from previous lines. We have numbered them consecutively with a B prefixed to each numeral. *Syll.* 2 is PM's *2.05.

The following lines are all obtained from detachment by the use of *Syll.* 2 as a derived transformation rule: B7, B10, B12, B16 and B19. First, B6 is obtained from *Perm* by the following substitutions in A3: $q \vee . p \vee r/q$. Then, by *Syll.* 2, from B5 and B6 we obtain B7 by iterated detachments. Second, B10 is obtained, by *Syll.* 2, through detachment from B8 and B9. Substituting: r/p, p/q in A3 to yield B9. Third, we get B12, by *Syll.* 2, through detachment, from B10 and B11. Substituting: $p \vee r/q$ in A2 to yield B11. Fourth, B16 is obtained, by *Syll.* 2, through detachment from B14 and B15. Substituting: $q \vee . p \vee r/p$ in A1 to yield B15. Last, we get B19, by *Syll.* 2, through detachments from B17 and B18. In PM the proof of *Comm.* (the Commutative Principle), viz:

$$*2.04.\ p . \supset . q \supset r : \supset : q . \supset . p \supset r$$

is dependent on *Asso.* And PM's proof of *Syll.* 2, viz:

$$*2.06.\ p \supset q : \supset : q \supset r . \supset . p \supset r$$

is proved with the aid of *Perm.* Hence Bernays's proof is circular.

University of Lagos

IF CONDITIONALS WERE NOT CONTRAPOSABLE . . .

By MICHAEL CLARK

> [C]ontraposition applies to no conditionals, of whatever type.
> (Peter Downing, ANALYSIS, 35.3, p. 88.)

PATIENT. I don't want an operation.
Surgeon. If I don't operate, you'll be dead by tomorrow.
P. I bet you I'll still be alive tomorrow.
S. Well, if you are, then I shall have operated on you.
P. Not necessarily. Granted that if you don't operate, I'll be dead. Untreated, I have no hope of survival. But during my long illness I have been reading up Downing's account of conditionals,[1] and it seems that I can quite consistently deny that if I live you'll have operated. You don't have any other reason for thinking that my survival requires an operation, do you?
S. No, I don't.
P. Then there is no reason why I should not look forward to being alive tomorrow untouched by the surgeon's knife.
S. I would try and persuade you to re-examine Downing's arguments, if I thought you had enough time left, because I can't help thinking that an account with such a curious consequence must be wrong somewhere. Harbour your forlorn hope, if you like. But don't forget that you have accepted my opinion that if you refuse an operation, you will die very shortly. Do you still refuse?
P. Yes, I do. I just cannot face a major operation: and so I shall prepare myself for bodily demise by reading McTaggart on the afterlife.

University of New Orleans and University of Nottingham

[1] See references in ANALYSIS, *loc. cit.*, and cf. his surgeon/patient example on pp. 87–8 of that paper.

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, R. G. Swinburne; J. H. Benson, Margaret A. Boden, Andrew Harrison, R. F. Holland, Hidé Ishiguro, J. Kemp, Bernard Mayo, D. H. Mellor, A. R. White, Peter Winch. This committee is responsible for appointing and advising the editor and for the general policy of the paper.

SUBSCRIPTIONS. The subscription to ANALYSIS for institutions is £5.00 (inland), £6.00 (overseas), $15.00 (U.S.A. and Canada); for individuals £4.00 (inland), £4.80 (overseas), $12.00 (U.S.A. and Canada). Each volume comprises four numbers, three of 48 pages and one of 64 pages, appearing within the academic year—in October, January, March and June. Orders should be sent to Basil Blackwell, 108 Cowley Road, Oxford OX4 1JF, or placed with any bookseller.

CONTRIBUTIONS. Articles submitted for publication should be addressed to Dr. C. J. F. Williams, 1 Fossefield Road, Midsomer Norton, Bath, BA3 4AS. Contributors are asked to note the following requirements.

Articles should **normally** not exceed 3,000 words in length. **Occasionally,** however, longer contributions can be accepted;

They must be typewritten in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

It is regretted that owing to increased postal charges it is no longer possible to return typescripts unless the following instructions are followed:

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgment is required, a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope **and international reply coupons** of the requisite value, whether for air or surface mail.

Galley proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts will be retained by the Editor on the assumption that authors have kept their own copies.

The copyright of articles printed in ANALYSIS remains the property of the author, but contributors are strongly advised, in their own interest, to consult the Editor before consenting to the reprinting of their articles.

PRINTED IN GREAT BRITAIN BY BURGESS & SON (ABINGDON) LTD., ABINGDON, OXFORDSHIRE

Vol. 36 No. 3 (New Series No. 171) March 197

ANALYSIS

Edited by
C. J. F. Williams

CONTENTS

BASIL BLACKWELL · ALFRED STREET · OXFORD

ANALYSIS 36:3 MARCH 1976

COHEN'S CRITICISM OF DUMMETT

By D. H. M. BROOKS

A WELL known riddle involves an island inhabited by two tribes: the members of one tribe always tell the truth, the members of the the other tribe always lie. The problem arises when a traveller going to a strange village arrives at a fork in the road. There is an islander there whose tribal affiliations are not known. What single question must the traveller ask to avoid misdirection? The solution involves a self-reflexive question and relies on the tribesman's second-level veracity or mendacity. The traveller must ask, for example, 'If asked, "Is the right-hand road the road to the village?" would you/a member of the other tribe say, "Yes"?'.

Mr. Cohen's doubts about Dummett's treatment of assertions (ANALYSIS 36.1, pp. 1–5) depend in part on our characterisation of the linguistic behaviour of the lying tribe, whom I will call the 'Mendacians'. The other tribe I will call the 'Veracians'. My characterisation of the Mendacians' behaviour so far has been misleading and is not germane to the point in question; for lying implies an intention to deceive, and a tribe which was always mendacious rather than always untruthful could not continue for long. For one thing, children could not be brought up to conform to the custom, since a Mendacian father by predicting successively 'square', 'invisible', 'carnivorous', etc., of a ball, rather than the pertinent predicate 'not-round', could so deceive his son as to render it impossible for him to learn the language of the island. (This possibility is also open to a rebellious Veracian, who could set back his son's development by successively predicating of a ball, 'not square', 'not the Taj Mahal', 'not a doll', etc.) Moreover, if the Mendacians go in for this kind of deceitfulness, one of their number could in some circumstances avoid being tricked into helpfulness by saying that he does not know, or by adding the "information" that the traveller must retrace his steps to get to the village.

The philosophically interesting case occurs when we do not impute deceitfulness to the Mendacians. They communicate well with each other, their children have no especial difficulty in learning to talk and their utterances are governed by all the considerations of relevance, helpfulness and "honesty" which govern the utterances of the Veracians (except perhaps the rebellious one). If this is the state of affairs on the island, there are, it seems prima facie, two ways of characterising it:

(1) Two completely different languages are spoken on the island, and it is completely irrelevant that exactly the same set of sentences is utterable in both languages

(2) The two tribes speak the same language, but the Mendacians have a convention of attempting to utter only false sentences, rather than true ones.

(We must formulate the convention thus to avoid the supernatural possibility of a tribe which *could* only utter false sentences.) We must note that the Mendacians need not bear the stigma of untruthfulness. If a linguist is the island's first visitor and he is looked after by the Veracians, whose language he learns, they will tell him about the untruthful Mendacians, and in his further explorations he will find that they have told him the truth about this untruthful tribe. If on the other hand he learns Mendacian first, he will be told about the untruthful Veracians and will discover that *they* are in fact untruthful. Unless the linguist is either very unorthodox, or committed to the belief that thesis (I) is the only correct characterisation of the state of affairs on the island, he will probably hold that the first tribe he encounters on the island has a convention of truthfulness, while the other tribe has a convention of untruthfulness. A very unorthodox linguist might assume that the first tribe he encounters has a convention of untruthfulness. The point is that there is no way of telling from a people's linguistic behaviour whether one should work from the hypothesis that they aim at making only true statements, or from the hypothesis that they try to make only false statements.

Cohen adheres to thesis (I) and argues that thesis (II) is incoherent, as it is 'no mere convention that speakers of a language mean to be taken as saying what is true' (ANALYSIS 36.1, p.4). He would hold that the Mendacians speak a different language in which the same signs have different sense, and that they try to make true statements. In support of his view he cites Wittgenstein's *Tractatus* 4.062,

> a proposition is true if things are as we use it to say that they are; and if by '*p*' we mean $\sim p$ and things are as we mean, then on the new interpretation '*p*' is true and not false.

Dummett raises the issue in discussing his contention that

> [t]he notions of sense and reference do not suffice for a complete account of language. If we know of a language only what sense the expressions which occur in it have, and thereby their reference, we know nothing which can tell us the significance of uttering an expression of this language: the point of doing so. (M. Dummett *Frege: Philosophy of Language*, p. 295.)

Cohen, however, holds that neither thesis entails that an account of language in terms of the sense and reference of its sentences (an SR

account) is inadequate. If we accept thesis (I) and hold that to speak a language at all a people must 'mean to be taken as saying what is true' (ANALYSIS 36.1, p.4), then the Mendacians speak a language in which they aim to make true statements, a language which is different from Veracian, and an SR account of this language will be completely adequate. We add nothing to the SR account by claiming that the Mendacians aim to make true statements, as to speak a language at all they must do this. On the other hand, according to Cohen, to accept thesis (II) and say that 'it makes no difference whether it be said that their intention is to be taken as saying what is true or as saying what is false is to make the notion of sense not inadequate but futile: we never could know the sense (nor therefore the truth-value) of a sentence of their language' (ibid., p.5). Cohen feels that an SR account of a language must be either quite adequate or futile.

I hold, on the contrary, that a clear counter-example can be given to this and that the problem of how to characterise the Mendacians' linguistic behaviour is not strictly relevant. This is because, even if we accept thesis (I) and agree with Cohen that the Veracians and Mendacians speak different languages, we are not forced to accept that an SR account gives an adequate account of a language.

Let us set our island story in the Twenty-First Century, so that our linguist can easily give a full SR account of Veracian, according to standard twenty-first-century practice which is based on some development of Tarski's theory of truth. The linguist realises that to formulate his SR account in terms of the notion of truth would lead him into philosophical difficulties. Accordingly he uses a more neutral notion, which also has the advantage of enabling him to give one SR account which serves for both Veracian and Mendacian. He is aware that Veracian linguistic practice divides the sentences utterable by Veracians into two classes: V, the class of sentences Veracians attempt to utter, and $\overline{V}$, the class of sentences they do not attempt to utter, and that Mendacians similarly divide Mendacian sentences into two classes, M, and $\overline{M}$. The situation on the island is such that, leaving aside problems raised by the possibility of truth-valueless sentences $V=\overline{M}$, and $M=\overline{V}$. So he formulates his SR account in terms of the predicates 'ε M' and 'ε V', pointing out that $V=\overline{M}$ and $M=\overline{V}$. He uses two predicates rather than one, so that the SR account will clearly cover both Veracian and Mendacian. His account will contain sentences such as these:

(i) $\ulcorner Fa \urcorner \varepsilon V$ iff a is round.
(ii) $\ulcorner Fa \urcorner \varepsilon M$ iff a is not round.
(iii) "$\ulcorner Fa \urcorner$ is verus" εV iff a is round.
(iv) "$\ulcorner Fa \urcorner$ is verus" εM iff a is not round.
(v) "$\ulcorner Fa \urcorner$ is falsus" εV iff a is not round.
(vi) "$\ulcorner Fa \urcorner$ is falsus" εM iff a is round.

(We must note here that 'verus' and 'falsus' mean 'true' and 'false' in both Veracian and Mendacian, though the extensions of 'verus' and 'falsus' in Veracian are the complements of their extensions in Mendacian. This is not the case with a predicate like '*F*', which, while its extension in Veracian is the complement of its extension in Mendacian, is not translated from both Veracian and Mendacian by the same English word. In Veracian '*F*' means 'round' while in Mendacian it means 'not round'.)

Let us suppose now that the linguist is shipwrecked on his return from the island, and all that survives the wreck is his SR account of Islandic, which does not contain any definition or intuitive account of the predicates 'ε *V*' and 'ε *M*', since the account itself provides a sufficient definition. Other linguists, who are unwilling to undertake the hazardous journey to the island, are surely in a position which Cohen would consider impossible. Not knowing the island situation they will be surprised by the SR account's containing so much apparently redundant material, but may account for this by any number of incorrect explanations, e.g. explanations ad hominem. The essential point is that they will not know which island language they have an SR account of. Some linguists may assume that *V* is the class of true sentences and translate '*F*' as 'round'. These linguists will have a complete account of Veracian, and will feel that the sentences involving the predicate '*M*' are redundant. Other linguists may assume that *M* is the class of true sentences, thus obtaining a complete account of Mendacian. The SR account must be supplemented with the information, unfortunately lost, that the Veracians attempt to utter only sentences belonging to *V*, while the Mendacians attempt to utter ony sentences belonging to *M*. The linguists SR account, qua SR account, is completely adequate, but is inadequate as an account of the languages of the island. This is Dummett's principal point.

Why then would Cohen deny that this could occur? It is difficult to find in his article any argument for this besides the one used in the quotation from the *Tractatus* given above. The rationale behind this argument is as follows: Cohen and Wittgenstein hold that 'it is no mere convention that the speakers of a language mean to be taken as saying what is true' (ibid., p. 4), or, as Dummett puts it in his paper 'Truth', 'it is part of the concept of truth that we aim at making true statements' (quoted ibid., p. 1). That is, if *S* is true, then *S* is a sentence we attempt to utter, while if *S* is false, it is a sentence we do not attempt to utter. Hence, if one attempts to utter only false sentences, one is attempting to utter only sentences one does not attempt to utter. This endeavour is logically doomed to failure. Since communicating by making false statements seems possible prima facie, we must describe such communication more precisely. We might say that if *S* is a true sentence then *S* is a

member of *the* class of sentences of language L, members of which L-speakers attempt to utter. The attempt to communicate by making only false statements is hence described as the attempt of an L-speaker to utter only sentences belonging to the class of sentences of L, members of which L-speakers do not attempt to utter. What vestigial paradoxicality remains can be removed by making the class of sentences of L, which L-speakers attempt to utter, a defining characteristic of L. The "L-speaker" is, hence, not speaking L at all, but a different language. This line of thought underlies the contention that our attempting to make true statements is part of the concept of truth.

Coupling this point with the view that an SR account of language must be given in terms of the notion of *truth* leads to Cohen's position, the belief that an SR account of a language must be a completely adequate account of the language. This I dispute, for, while it may not be mere convention that L-speakers attempt to utter *true* sentences, it *is* mere convention that they attempt to utter the class of sentences that they do rather than the other class of sentences in the language. Hence an SR account which merely distinguishes these two classes, without specifying which of these two classes the language speakers attempt to utter, is an inadequate account of the language. This is shown by the fact that *one* SR account—such as the one I have outlined—can be common to *two* languages. This is possible because the concept of truth, as Cohen conceives it, need not be the guiding principle of an SR account.

The chief reason for holding that an SR account of a language is inadequate is, of course, that it enables us to deal with assertion on a par with other types of linguistic force, such as imperative or interrogative force. For Cohen assertion would disappear into the SR account and it would be difficult to find anything common to both 'The door is open' and 'Open the door'.

University of Natal, Pietermaritzburg

ON *EXPRESSING* FUNCTIONS AND PROPERTIES

By PETER LONG

A FUNCTIONAL expression, like any other expression, is a quotable sign. Thus 'log' and '$\times$' are functional expressions, whilst '3' and 'π' are not. Now a function that is expressed by a functional expression is expressed by that expression's being combined with one or more names (singular terms) or variables. The logarithmic function is expressed by the sign 'log' being combined with a numeral or variable and the multiplication function is expressed by the sign '$\times$' being combined with two numerals or variables (or a numeral and a variable). Since a function that is expressed by a functional expression is expressed *by* that expression's being combined with one or more names or variables, it is expressed *in* any sign that is so formed: the logarithmic function is expressed in 'log 5' and 'log x' and the multiplication function is expressed in '2×3', '$x \times y$' and '$2 \times y$'.

It may seem strange to say that a function that is expressed by a functional expression is expressed by that expression's being combined with a name or variable. We may be inclined to say '*Either* the logarithmic function is expressed by "log" *or* by "log" being combined with a numeral—we can't say both!' But we can, and to think otherwise bespeaks a failure of ear for two different uses of the verb 'express'. When we say that a certain function is expressed by a functional expression we do not mean that it can be expressed by simply writing that expression down unaccompanied by the name of an argument, or by the variable 'x' or the word 'something'. That would be absurd. We mean that it is *identified* by an expression, so that the complex signs in which it is expressed contain a piece of language which *shows* that it is expressed in them. Thus the logarithmic function is identified by the sign 'log': it is from the presence of this sign that we recognize the function expressed in 'log 5' and 'log x' to be the logarithmic function. When, on the other hand, we write '5^2' and '3^7', then it is not from the presence of any sign that we recognize that the function expressed in these two complex signs is the exponential function; but if we were to *write* '5 to the power 2' and '3 to the power 7', which is how we read '5^2' and '3^7', then we should be employing a notation in which the exponential function is identified by an expression. So we could point up the difference between the two uses of 'express' by saying 'The exponential function as it is expressed in "5 to the power 2" is expressed by the sign "to the power ", but as it is expressed in "5^2" it is not expressed by any sign'.

But now why should we use the word 'express' twice over and say that the logarithmic function is 'expressed' by 'log' being combined with a numeral, thus seeming to invite confusion with that use of

'express' in which a functional expression expresses a function? Why should we not say instead that the logarithmic function is 'designated' or 'symbolized' by 'log' being combined with a numeral? The answer seems to be that we employ the word 'express' rather than 'designate' or 'symbolize' because what follows the word 'by' in the above two sentences is not the quotation or description of a sign or symbol, but the description of a fact. We can say that such-and-such a fact 'expresses' a function, but we can hardly say that it 'designates' or 'symbolizes' a function.

It is a convention that the exponential function is expressed by one numeral being inscribed on the right shoulder of another, and it is because we are acquainted with this convention that we recognize what function is expressed in '5^2' and '3^7'. But it is not because we are acquainted with a convention that we recognize what function is expressed in 'log 5', but because, as we should say, we know what function 'log' expresses. Thus it is in virtue of a convention that '5^2' and '3^7' are complex signs and not mere complexes of signs, but there is no convention in virtue of which 'log 5' is a complex sign and not a mere complex of signs. It is of course a convention that 'log' is used as a functional expression, as it is a convention that '5' is a numerical expression; but the point is that there is no convention in virtue of which a complex sign whose components are a *functional expression* and a numeral is a complex sign.

It is because there is no convention in virtue of which 'log 5' or 'log x' are complex signs that we say that the logarithmic function is expressed by the sign 'log' being *combined* with a numeral or variable, for 'to combine one sign with another' does not *mean* 'to put them into some relation with one another'. We could not say that the function expressed in '5^2' and '3^7' was expressed by one numeral being combined with another. Indeed two numerals will not 'combine' to form a third sign. Where we express a function f by combining a sign 'S' with a name or names, it is clear that 'S' itself must show that the function f is being expressed and so must be a functional sign.

Of course we cannot combine two written signs without relating them spatially and in combining 'log' with a numeral or variable we adopt the practice of putting the functional expression first in the linear ordering. With the functional sign '!', on the other hand, we adopt the practice of putting the numeral first and write, not '!5', but '5!', though, significantly enough, we read this as 'factorial five' and not 'five factorial'. We can read it so because it is the fact that '!' and '5' are combined with one another that expresses the function and not the fact that '5' precedes '!' in the linear ordering. In 'log 5' and '5!' the ordering of the component signs relates only to the particular way in which they happen to be produced and is not what expresses. Again, the function that is

expressed in '2 — 3' and '3 — 2' is expressed by the minus sign's being combined with '2' and '3', and if we were to introduce the functional sign first and wrote '— 2,3' instead of '2 — 3', as is the practice in some logic texts, we should not be exploiting a different convention for expressing the subtraction function. For once again, we are not exploiting a convention at all when we express a function *f* by combining a sign or expression for *f* with one or more names

A functional sign or expression need not be a unitary one like 'log' or '×'. The functional expression contained in 'the midpoint between A and B', for example, can only be defined by using brackets as letters to indicate where the names that are combined with it are to be put. Nothing of logical significance hangs on this fact and we could express the same function by combining the typographically simple expression 'midpoint' with 'A' and 'B' and writing 'A midpoint B' or 'midpoint A, B'. So if we say that we express the function that is expressed in 'the midpoint between A and B' by combining 'the midpoint between ξ and ζ' with 'A' and 'B', we must distinguish the role of the Greek letters here from the role they have when they are introduced, not to define a functional expression, but to define a function. Thus Frege would refer to the function that is expressed in '2 — 2' and '3 — 3', but that is not expressed in '2 — 3', as 'the function ξ — ξ', and here the letter 'ξ' is *not* used to enable us to recognize the functional sign that is contained in '2 — 2' and '3 — 3' but not in '2 — 3', for there is none. No *sign* shows that the function ξ — ξ is expressed in '2 — 2'. Here the Greek letter is used to enable us to recognize the function that is being referred to. Therefore, if we speak, as Frege would of 'ξ — ξ' as 'an expression for a function' and say that '2 — 2' is the result of 'completing' the expression 'ξ — ξ' with the numeral '2', this does not mean that it is the result of 'completing' a functional expression—a piece of language that shows what function is being expressed—with the numeral '2'.

We can speak of *the* function that is expressed in '2 — 3' and in '3 — 2', but not of *the* function that is expressed in '2 — 3'. For of course three different functions are expressed in '2 — 3', namely the minus function, the function ξ — 3 and the function 2 — ξ. Now we cannot omit the Greek letter here and say that, in addition to the minus function, the functions — 3 and 2 — are expressed in '2 — 3'. If we did this, we should not know what functions were being referred to. But this is not because '2 — 3 does not contain unitary expressions for the functions ξ — 3 and 2 — ξ, but because it does not contain expressions for these functions at all. These functions are expressed in it, but it is not through the presence of any sign that we recognize that they are expressed in it, as it is through the presence of a sign that we recognize that the minus function is expressed in it. If we could take the last two signs in '2 — 3' as constituting an expression for a function of one argument, then the notation

that is canonical in certain logic text-books in which the sign or signs for the argument(s) of a function are written after the functional expression would be incoherent. When we thought of '— 3,2' as combining the sign ' — ' with two numerals, we should have to understand it as meaning '3 — 2' (which it does), but when we thought of it as combining the 'expression' '— 3' with '2', we should have to understand it on the analogy of 'log 5', in which case there would be no difference in sense between '— 3,2' and '2 — 3'.

We can see, in the light of this, why Ramsey should have thought it incomprehensible that a proposition such as 'Socrates loves Plato' could be analysed in three different ways: as asserting that the relation *loves* relates Socrates to Plato, as asserting that Socrates has the relational property of loving Plato, and finally as asserting that Plato has the relational property that Socrates loves him[1]. The mistake that Ramsey made was to suppose that it could be analysed in either of the last two ways only if it contained, which it does not, an *expression for* a relational property. Two different relational properties are indeed expressed in this proposition, but the proposition contains no expression for a property, relational or otherwise. Ramsey was discussing the problem of universals and if we call both properties and relations 'universals', then we can say that the only expression for a universal that the proposition contains is the expression 'loves'. On the other hand, if we may say that a person is a Christian if and only if he loves Christ, then the proposition 'Luther is a Christian' does contain an expression for a relational property, but then it does not contain an expression for a relation.

Now we referred just now to Frege and said that he would speak of '2 — 2' as the result of completing 'ξ — ξ' with the numeral '2'. But what here undergoes completion, and is thus itself incomplete, is not an expression, not a piece of language. So we are led to ask whether there is any reason for calling a functional expression proper, as opposed to a name, 'incomplete'. Certainly a functional expression has a role incommensurate with that of name, but why should this be such a reason? Since the role of a functional expression is to show that such-and-such a function is expressed by combining it with a name or names, it can only express a function when combined with one or more names (or variables); but then, it may be said, a name can only represent an object when it occurs in the context of a proposition or, at least, when the context of its occurrence enables us to supply a propositional setting for it. This consideration does not mean that the roles of names and functional expressions are not radically different, but it might seem to preclude us from contrasting them as respectively 'complete' and 'incomplete'.

In fact it does not. In order to see this, let us consider the difference between the two complex signs '5^2' and '5 to the power 2'. The function

[1] *The Foundations of Mathematics* (Routledge & Kegan Paul, 1954) pp. 117—119.

to the power of is expressed in both these signs, but in the first it is expressed by putting '5' and '2' in a certain relation to one another and in the second it is expressed by combining them with the expression 'to the power'. Suppose we now ask how it is possible to express the function in these two different ways. How can what is expressed by relating two numerals in a certain way be also expressed by combining them with a third expression? The answer seems clearly to be: Because in expressing the function *to the power of* by combining '5' and '2' with 'to the power' we are still expressing the function by putting the numerals in a certain relation to one another, only the relation is now an indirect and unspecific relation. In the complex sign '5 to the power 2' (or the complex sign 'to the power 5,2') the numerals are related to one another through being related spatially to the inscription which embodies the functional expression, and it is this indirect and unspecific relation by which the exponential function is expressed in '5 to the power 2'. But now if the inscription 'to the power' thus defines the relation by which the function is here expressed, then this must, so to speak, be written into the expression which the inscription embodies. The expression 'to the power' must therefore contain, as we may put it, the form of the signs with which it combines to form complex signs in which the exponential function is expressed: it must contain the form of two names. For this reason we may find it natural to call the expression 'to the power' incomplete, and in speaking about it to accompany it with two brackets or letters just in order to show that it is an expression for a function (of two arguments).

Of course what we have said here about the expression 'to the power' applies *mutatis mutandis* to expressions for functions of one argument: thus the function expressed in 'log 5' is expressed by the unspecific relational property that a numeral has when it is combined with 'log'.

We have been concerned in this paper with expressions for what Frege would call 'first-level functions', and we can see now that *in the sense in which* there are 'expressions for' first level-functions, there are no expressions for what Frege would call 'second-level functions'. For just because expressions such as 'log' and 'to the power' have the role they do, there can be no expressions which contain the form of expressions of this class as expressions of this class contain the form of names. Now it might seem to follow from this that we are at loggerheads with Frege, since he held that expressions for functions, of whatever level, were incomplete. But this really follows only if his use of 'functional sign' and 'expression for a function' accords with ours. And this seems not to be the case. For when he uses these terms he is generally, if not always, using them in a sense different from the normal one. Thus when he writes 'the expression for a function must always show one or more places that are intended to be filled up with the sign of the argument'[1], he is not speaking of a

[1] *Translations from the Philosophical Writings of Gottlob Frege* (Blackwell, 1960), p.25.

functional expression proper at all. As he is here using 'expression for a function' *any* complex sign contains an expression for a function provided it is one in the utterance of which a function is expressed. Thus '5^2' no less than 'log 5' contains such an expression, and in the corresponding use of 'functional sign' '2 — 2' and '3 — 3' contain a functional sign that is not contained in '2 — 3'. So understood, a functional sign or expression is not really a sign or expression at all; it is not a piece of language that can combine with other signs to form a complex sign. Therefore although Frege would say that 'log 5' is composed of 'log ξ' and '5', the former being incomplete, he need not mean that it is composed of the two *expressions* 'log' and '5', the Greek letter being attached to the former merely in order to show that it is a functional expression. He may mean rather that 'log 5' results from the schema 'log ξ' by replacing 'ξ' by '5', where the Greek letter is meant to indicate that this schema is one in the exemplification of which a first-level function of one argument is expressed.

What we have said of functional expressions and expressing functions applies unchanged to predicates (expressions for properties of objects) and relational expressions and to expressing properties and relations. Thus the sense in which a predicate or relational expression 'expresses' a property or relation is the same as that in which a functional expression 'expresses' a function: the expression 'bald' shows that the property expressed in 'Socrates is bald' and 'x is bald' is the property *bald*, and the sign 'next to' shows that the relation expressed in 'Socrates is next to Plato' or 'x is next to y' is the relation *next to*. That predicates and relational expressions are thus comparable in role to functional expressions does not mean that properties and relations are functions, but it does mean that the relation between the concept of a property or relation and the concept of a function is a peculiarly intimate one.

If the sense in which predicates and relational expressions 'express' properties and relations is the same as that in which functional expressions 'express' functions, then the role of such expressions has been radically misconceived by most philosophers. Just because we call 'bald' an expression for a property and 'next to' an expression for a relation, most philosophers have assumed that it is the presence of the word 'bald' that *constitutes* 'Socrates is bald' a proposition of the subject-predicate (predicative) form, and the presence of the words 'next to' that *constitutes* 'Socrates is next to Plato' a proposition of the relational form—as, for example, the presence of the name 'Socrates' *does* constitute 'Socrates is bald' a proposition about a person. But the truth is, not that predicates and relational expressions are constitutive of the predicative and relational form, but that they make it *verbally* manifest that the propositions, or clauses, in which they occur are of these forms. In a notation that employed the convention of expressing the relation *next to* by writing a name or

variable next to a name or variable, the formation 'Socrates Plato' would be a proposition in which the relational form was not verbally manifest.

Philosophers have erred in not taking seriously enough the simple fact that we call predicates 'expressions for properties' and relational expressions 'expressions for relations' because they are signs that 'express' properties and relations. On the other hand, we do not, for instance, call a personal proper name 'an expression for a person' because it is a sign that expresses a person! If we are struck by this, as we should be, and go on to ask 'Why is it that we speak of a predicate or relational expression as "expressing" a property or relation?' we are within reach of the correct conception of the role of predicates and relational expressions. We may then come to see that it is essential to a property or relation, as opposed to what Frege calls an object, that the only thing that an expression for them *can* do is to *express* them—that is, to make it verbally manifest that the propositions in which they occur are of the predicative or relational form. If we may be allowed a paradox, it is just because a property or relation is not expressed by a piece of language, but by a fact, that a piece of language can only express the predicative or relational form. Here of course we are playing on the double use of 'express' to which we drew attention at the beginning of this paper.

Properties and relations, like functions, are expressed by properties of, and relations between, signs. Thus to say that the property expressed in 'Socrates is bald' is expressed by the name's being combined with 'bald' is to say that it is expressed by the name's having the property of being combined with 'bald'. Again, to say that the relation expressed in 'Socrates is next to Plato' is expressed by the two names being combined with 'next to' is to say that it is expressed by their being related as two names are related when they are combined with the sign 'next to'. Confining ourselves to the written language, this relation is the indirect and unspecific relation that two names have to one another when they are related to the inscription 'next to' in the way that two names are related to it when they are combined with the sign that the inscription embodies. If we call this relation 'R', then it will be true to say that whereas the relation expressed in 'Socrates is next to Plato' is expressed by the names standing in the relation R to one another, the relation expressed in 'Socrates Plato', where this is written in the notation we imagined, is expressed by the names being next to one another. But now, as there is no convention in virtue of which the sign '2×3' is a complex sign, so there is no convention in virtue of which the sign 'Socrates is next to Plato' is a proposition: as there is no convention in virtue of which a complex sign that consists of two numerals and a functional expression is a complex sign, so there is no convention in virtue of which a proposition that consists of two names and a relational

expression is a proposition. And yet it is a convention that a relation is expressed by two names being in the relation R to one another, just as it is a convention of our imagined notation that a relation is expressed by two names being in the relation *next to* to one another! But there is no inconsistency here, as one might at first suppose. For the convention that a relation is expressed by two names being in the relation R to one another just *is* the convention that the inscription 'next to' is the inscription of a relational expression! Thus an understanding of the convention that a certain inscription is the inscription of a relational expression itself reveals that such an expression manifests, but does not contain, the form of the propositions in which it occurs. And of course the same is true of the convention that a certain inscription is the inscription of a predicate.

University of Leeds

A PUZZLE ABOUT DREAMING

By J. F. M. HUNTER

STRANGE things happen in dreams. A walk in the woods turns into a tense scene in a doctor's office; a chat about the weather is felt as being dramatic and intensely difficult; one flies like a bird, or composes poetry with the utter facility with which it might be recited from memory. It is not such peculiarities of dreams that I wish to dwell on.

Consider a rather different oddity. I dream of an engrossing conversation. On waking I tell the dream, not with the wide-eyed fascination with which I might tell of having flown like a bird, but just as if it were a conversation of comparable interest that had occurred in waking life. There seems nothing fabulous about the dream incident; and yet when I remember it quite precisely I realize that the person with whom I talked had a rather peculiar presence in the dream: sometimes there was just his voice, sometimes a wistful or a puzzled expression and nothing more, sometimes just a gesture of his hands as he spoke. The fragmentary nature of his embodiment in the dream was not due, as it sometimes is in waking life, to my having closed my eyes here, focused my attention on his hands there, and so on. There *was* nothing more, at a certain point, than hands gesturing; and yet while there was not at that moment a human body along with them, there was not the absence of a body either: they were not seen as hands moving eerily before my eyes as if to gesture. Had I not remembered the dream precisely I would, with no sense of wonderment, have described this point in the dream simply as his having made an impatient gesture with his hands as he said this.

Dreams can be unspecific as to whether an incident occurred indoors or outside, how a person was dressed, whether he was standing, sitting or walking, and so on; but if nothing is given as to someone's facial expression at some point, it is not that there is a face, but it is neither smiling, frowning nor anything else, nor does a dream figure appear at such junctures as lacking a face. There is not a hole where the face would have been, or a body without a head.

If I dream of walking along a street and seeing a house with an interesting roof, the house itself may be neither brick nor frame nor any other construction, and nothing may be given as to the placing of its windows and doors. Yet it is not just a roof hanging in the air, nor are its walls of unidentifiable materials, or unbroken by windows and doors.

Real conversations and roof-sightings occur in a plenum of detail. Conversations happen either indoors or outdoors; the participants at any given time are either standing, sitting, walking or lying down, and

are smiling, frowning or looking bored. It is either true or false that they have moles on their chins or wallets in their pockets. One may not notice the detail, or may not remember it; but in real life there is then necessarily something missed or forgotten. In dreams this requirement is lifted, and if we cannot right away say afterwards what someone was wearing or doing, there may or may not be an answer: it may come back to us that she was striding along purposefully as she said this, but also we may remember that the dream included nothing as to what she was doing.

Insofar as a dream includes this peculiarity, no real life pictorial representation of it seems possible. If she smiled as we talked, but nothing appeared as to her posture or dress, there was not just a face or a head, floating eerily before our eyes; but on the other hand it is impossible to depict a body that is neither still nor moving, neither seated nor standing. Yet the smile itself was there. We sometimes remember it vividly, and could draw it quite precisely.

This feature of dreams can strike us as very odd when we think about it, but it is not odd in the way dreams can be odd. We would not say, in astonished tones, even when we remember just how it was, that we had a conversation, sometimes with a face, sometimes with a fist shaking menacingly, sometimes with a voice out of nowhere—because that is not how it was. There was not a face in a visual field, surrounded either by nothing or by background trees and fields or walls and curtains. (That might happen in a dream, but it is not the case we are considering.) Nor is the visual field wholly occupied by the face, as it sometimes is in a cinema close-up, looming very large. It is the face of a person, in some ordinary scene, a room or a garden—only the rest of the person and his surroundings are (sometimes) not given. Further, it is not just that they are given at certain junctures, as when our attention is focused on something else. They are sometimes not given *at all.*

We have a model of this peculiar fact in an average novel. Chapter Seven opens with Paul and Henrietta intensely discussing the state of their marriage. Nothing is said as to how the conversation got started, where or when it occurred, whether they smoked or drank coffee, how they were dressed or whether they smiled or frowned as they talked. We may never have been told whether Henrietta is red-haired, tall or English, whether Paul suffers from stomach ulcers, or what he does for a living. Yet we do not get goose-bumps reading it, seeing it as a conversation occurring nowhere and nowhen, between disembodied voices having neither this nor that tonal quality.

If it is now clear that there is something very odd indeed about some dreams we might, if we are not inclined to insist that everything in life should conform to a few familiar patterns, simply list this amongst the fascinating surprises in store for the perceptive observer. There is not

yet anything puzzling about the fact that, contrary to what we might have expected, dreams should have this extraordinary character. A puzzle may emerge, however, if we ask how it is that, on the basis of the incomplete and peculiar fragments that are given in a dream, we so matter-of-factly report it as if it were a fully-articulated incident or sequence—being chased by a tiger, seeing an interesting roof, having an absorbing conversation—rather than perhaps as 'racing through nothingness pursued by a tiger's head', 'seeing a roof on a house that neither did nor did not have windows and doors', or 'having a conversation with a voice out of nowhere'.

Do we work with the fragments and fill in the details? Or do we not bother with the details assuming that we know at least *what kind* of larger picture the fragments would be parts of? Or is it somehow not at all on the basis of what we experience that we say how the dream went?

Clearly we do not fill in any further details. We might, as an indication of the sort of roof it was, say that it would have gone nicely with shuttered windows and a fanlight over the door; but this would be an indirect description of the *roof*. We would be fabricating if we flatly described the windows and doors that way.

Nor do we say 'I saw a roof which I took to be the roof of a house of some description'—as we might if we were on a Hollywood set and saw a roof rising over a tall hedge. Here there would be the possibility that it was a mock-up, the roof really being supported by metal scaffolding. We make no guesses or surmises. It makes no more sense to entertain the possibility of being wrong, or right either, about parts of the structure we did not "see" than to make guesses as to whether it was raining when the conversation at the beginning of Chapter Seven occurred if the weather was nowhere mentioned in the story.

Let us therefore explore the possibility that it is somehow not on the basis of what we experience while asleep that we tell our dreams. This, if it could be made intelligible, would resolve the puzzle, because it would then no longer come out as a fabrication to say, for example, 'I was chased by a tiger', with its implication of a plenum of detail, when the whole of what we experienced was perhaps a sense of struggle and terror, with occasional glimpses of burning eyes and snapping jaws.

Let us entertain the idea of a rather unusual raconteur: a very inventive fellow who not only can compose elaborate stories as he goes along, but sometimes does this for his own gratification, and when he does does not need to articulate the story he is composing, as he would of course have to do if he were entertaining someone else with it. He knows how his story runs and he could lay it out in rich detail, but when alone he does not go to that trouble. Let us imagine further that to enrich his imaginative endeavours he has invented a device with which, as a story unfolds, he can illustrate its features with representations, of

woodland scenes, screeching brakes, wistful smiles—many of them schematically or impressionistically done.

If we were to watch him using this device, we could generally make little sense of what we saw. Without knowing what story they illustrated we would find the pictures too impressionistic and their sequence consistent with any number of different interpretations. Since we are supposing that he would not narrate the story for us, we would be left wildly guessing as to its drift.

However he would have none of the difficulty we had; nor would the task that was difficult for us be easy for him. For him there would be no task at all. Being its author he would know what the story was. He would not have a moment's puzzlement over why the switch from the rose window to the careering ambulance, or whether that object racing through terrain that was neither town nor country was indeed an ambulance. These things would all meld together with the story he could, but did not, articulate, and leave him in a position afterwards to recount the tale without in any way guessing, or interpreting, what had happened.

What status might we assign to this idea of a raconteur? That is to say, what, if anything, corresponds to it in the human constitution? His mode of functioning is consistent with certain facts, chiefly the fact that although what comes before the mind when we dream is too fragmentary and curious (in the way explained earlier) to provide a basis for the way we tell our dreams, it does not seem to us that we interpret what has come before our minds or construct the story we tell. But we would certainly not want to add the raconteur to the *homunculi* with which the human psyche is sometimes populated.

The raconteur idea is primarily offered as an alternative to another idea that equally has no counterpart in the human constitution, the idea that dream tellings are a special kind of eye-witness report, that just as what we remember and tell about a fire or a coronation is all derived from the passing show of events, so also are our dream tellings, only in their case the passing show is inward.

The eyewitness picture is consistent with certain facts, chiefly the fact that dream experiences have the obstinate, independent character of things seen. They come unbidden, behave as they please and not as we do, and can have most unexpected properties. On the other hand, a *plain* eyewitness report of a dream would have a peculiar character. Instead of 'Martha and I were sitting in her kitchen talking', the eyewitness report would perhaps run 'There was a sketchy appearance of a room, which I would not have recognized, but which in spite of that I quite routinely took to be Martha's kitchen, and glimpses of a person who did not look like anyone I know, but whose unfamiliarity did not in the least strike me, and who seemed to me to be Martha . . .'.

In the raconteur and the eyewitness we have two different perspectives on the facts. It may be that properly we should have no view *of that type* about dreaming, but our problem derived from assuming the eyewitness picture; and even if all such pictures are fantasies, it is at least clear now that a different and more interesting fiction is available to us—that we are not bound by default of other candidates to suppose that truthful dream tellings are derived from the passing (inner) show.

It should be noted that scepticism about private experiences has played no role in these deliberations. It is not the fact that we do not see wistful smiles and monster rabbits in our dreams, but the fact that we do, and that they can have a peculiar character, that is puzzling. It is the fact that dream experiences, when we remember them quite precisely, often seem altogether incommensurate with the dream we conceive ourselves to have had that generates our problem.

We, may, as suggested just now, be trading in fictions here, but if so they are fictions we seem strongly driven to employ; and if one fantasy distorts our vision while another helps us to see clearly, it may not matter whether it is all a magic journey. It may be unapologetically suggested in conclusion, therefore, that there are interesting possibilities of applying the raconteur concept more widely.

We are apt to suppose, for example, that when we say what we have been thinking, the mechanics of it are like those of an eyewitness report: something happened from which what we later say is derived. It can then seem very puzzling, when we recall the character of what did happen, how it is that we are able to read in it the thoughts we so confidently say we had. One toys with such suppositions as that there is a language of thought, different from all spoken or written languages, and that we have a remarkable facility for translating it into English, German or Swahili. In this and other conjectures we might consider we are trying to make the facts fit an essentially eyewitness report picture of the mechanics of saying what one has been thinking.

We are entirely relieved of this quandary by the raconteur picture, the central feature of which is that we can be ready to do something, or on the point of doing it, without yet having done anything whatever, whether publicly or privately, that shows *what* we are on the point of doing. Accordingly, having a thought can be regarded as its coming about that one is ready to do something; and later expressing the thought, whether publicly or privately, can be regarded as doing what one had been on the point of doing. The expression of the thought is not based on noticing one's state at the time it occurred, or *based on* anything else, but is a matter of, so to speak, letting oneself go, doing something that one now *can* do. The experiences that some people tell us they have when they think, of talking to themselves, of picturing, and so on, need not be disbelieved: they can be regarded either on the analogy of

the raconteur's device for illustrating the story he could tell, or as initial expressions of the thought he has had.

The idea here of being ready to do something is a cousin of the concept of an ability: it is having an ability suddenly, and perhaps for a short time. Having an ability is not an experience. I can play chess, explain the ontological argument, recite various passages from *Hamlet*, and so on. I have these abilities day in, day out, sleeping and waking. I can experience my playing chess, but not my ability to play; and similarly I do not experience the ability to make an interesting suggestion, when suddenly a thought occurs to me. I know what my suggestion is, but not by introspection. To say that I know what it is is just to say that I can go right on from having the thought to expressing it.

It may seem astonishing and mysterious that we should know what we are thinking without there being anything to show what it is, but perhaps it is remarkable only because of the grammatical similarity between 'I saw such and such' and 'I thought such and such'. For it to be true that I saw such and such, something of that description must have occurred. We therefore expect that similarly something must have occurred, properly describable as 'the thought that such and such'. Certainly I am a deceiver if I say I was thinking so and so when I was not; but that is the same grammatical illusion in another guise, reinforcing our expectation that thinking will be a describable process and our astonishment on finding that it is not.

University of Toronto

MEANING, REFERENCE AND TENSE

By CLIFFORD E. WILLIAMS

RECENT literature on such topics as the status of temporal becoming, subjective versus objective time, thing-ontologies versus event-ontologies and perspicuous languages has often contained discussions of the way tensed and tenseless sentences should be distinguished. One important view is that tenseless sentences have the same sense every time they are used, whereas tensed sentences have different senses every time they are used because each tensed sentence is translatable into a different tenseless sentence at each different time. In a recent article, Stephen E. Braude claims that this way of distinguishing between the two kinds of sentences is mistaken because 'it does not square with the brute facts of ordinary discourse. Nonsimultaneous replicas of tensed sentences *can* have the same sense'.[1]

For ease of reference, I shall state Braude's 'brute fact' as follows

> (B) Nonsimultaneous replicas of tensed sentences have the same sense.

The plausibility of (B) depends on which theory of meaning is true. Unfortunately, Braude does not tell us which theory of meaning he adopts. Nor does he argue for (B) and defend it against objections—it is just a brute fact which no reasonable person can doubt. In what follows, I shall discuss (B) with respect to three important theories of meaning: the rules of use theory, the content theory and the reference theory.

According to the rules of use theory of meaning, the meaning of a word or sentence is the set of rules governing its use. If the rules of use theory is true, then (B) is certainly true—nonsimultaneous replicas of tensed verbs, as of other indexical expressions, are governed by the same rules of use. It is not necessary to issue a new dictionary every moment so as to keep up with the latest use of such words as 'was', ['now',] 'had', 'went', 'is going', and the like. We all know perfectly well how to use these words at many different times, but we surely wouldn't if they were used in a different way at each different time. We ought, therefore, to accept (B) when it is given a rules of use interpretation.

However, (B) is a good deal less plausible when it is given either a content or a reference interpretation—nonsimultaneous replicas of tensed sentences do not obviously express the same content or refer to the same states of affairs. Users of the English language appear to think that 'now' is used to refer to the time it is uttered or written—ask, if you

[1] *The Philosophical Review*, LXXXII (April, 1973), p. 203. I ignore the 'can' in what follows so as to simplify matters. No essential distortion results from doing this.

dare, the next ten people you see, 'What are you referring to when you use "now" in a sentence like "It is raining now"?' One wonders how a tensed sentence, such as

(1) Nathan is reading a book now

can express the same content, or refer to the same state of affairs, at different times if those who typically use (1) would say they are referring to a different time at each different time they utter (1). Moreover, it is commonly thought that sentences like 'I am six feet tall' and 'This is a maple' express different contents and refer to different things when uttered by different people or near different trees. (1) seems to operate in the same way that these sentences operate. All of them can have different truth-values when produced in different circumstances, and in order to know what their truth-values are, we have to know what circumstances they are produced in. So it would seem that (1) too expresses different contents and refers to different states of affairs at different times. Therefore (B) is questionable if either the content or reference theory is true. Indeed, if the considerations in this paragraph are correct, the claims that non-simultaneous replicas of tensed sentences express different contents and refer to different states of affairs have a certain degree of plausibility.

Since (B) seems so obviously true when interpreted according to the rules of use theory of meaning, but questionable when interpreted according to the content or reference theory of meaning, we need to ask whether anyone has ever claimed that tensed sentences have different rules of use each different time they are used. Could it be that those who distinguish between tensed and tenseless sentences in the way Braude objects to are really asserting, more plausibly, that tensed sentences express different contents or refer to different states of affairs each time they are used? Let us look at what is actually said by those who according to Braude adopt the position he argues against. Nelson Goodman, whose position Braude says he discusses (p. 201), writes that

> the 'now' in question [the 937th word uttered by George Washington in 1776] is translated by any 'The period referred to by the 937th word uttered by George Washington in 1776' .... Or we may seek a translation that contains no name of the indicator itself, but rather another name for what the indicator names. Thus a certain 'here' is translated by any 'Philadelphia', and a certain 'ran' is translated by any 'runs [tenseless] on January 7, 1948, at noon E.S.T.'[1]

In this passage Goodman is claiming that tensed sentences can be translated into tenseless sentences. A replica of (1), Goodman would say, can be translated into

(2) Nathan reads [tenseless] a book at t_1,

[1] *The Structure of Appearance*, second edition (Indianapolis: Bobbs-Merrill, 1966), p. 369.

where t_1 is the time of the production of the replica of (1), and another replica of (1) can be translated into

(3) Nathan reads [tenseless] a book at t_2,

where t_2 is the time of production of the new replica of (1), etc. As Braude points out (p. 202), it is because of this translatability of tensed sentences into tenseless sentences that Goodman believes that tensed sentences have different senses each different time they are used. That means we can find out which version of (B) he denies by finding out what kind of translatability he adopts. Unfortunately, Goodman does not tell us in this passage, nor in the context of this passage, whether he is using 'translated' to mean 'translated without change of meaning' or 'translated without change of reference'. In contemporary philosophical literature 'translated' is usually used as short for 'translated without change of meaning' and meaning is usually construed in terms of rules of use. Goodman, however, says something that might lead one to suspect that in this context he meant 'translated' to be short for 'translated without change of reference'. In the passage quoted above he writes that one 'may seek a translation that contains no name of the indicator itself, but rather another name for what the indicator names'. From this remark one might think that Goodman is claiming that the referent of a certain 'now' (i.e., what 'now' names) is the same as the referent of any 'The period . . . ' (i.e., what any 'The period . . . ' names). It is possible, therefore, perhaps even probable, that Goodman is espousing a reference translatability instead of, or in addition to, a rules of use translatability. Consequently, it is possible, perhaps probable, that Goodman believes only that tensed sentences refer to different states of affairs each different time they are used.

W. V. O. Quine is also mentioned by Braude as holding a Goodmanian-type position. In a passage that has sometimes been quoted in the philosophical literature on tense Quine claims that

> logical analysis is facilitated by requiring rather that each statement be true once and for all or false once and for all, independently of time. This can be effected by rendering verbs tenseless and then resorting to explicit chronological descriptions when need arises for distinctions of time. The sentence, 'The Nazis will annex Bohemia', uttered as true on May 9, 1936, corresponds to the statement, 'The Nazis annex [tenseless] Bohemia after May 9, 1936'; and this statement is true once and for all, regardless of date of utterance.[1]

In this passage Quine is clearly and unequivocally talking about the rules of use of sentences. He says that verbs can be rendered tenseless, i.e., that tensed verbs can be eliminated and replaced with explicit chronological descriptions, and that one sentence (a tensed one) corresponds to

[1] *Elementary Logic*, revised edition (New York: Harper, 1956), p. 6.

another one (a tenseless one). He does not say that tensed verbs can be eliminated and replaced with explicit chronological descriptions without a change of rules of use. Nor does he indicate that the 'correspondence' of the two sentences is an identity of rules of use. Furthermore, an examination of the context of the quotation from Quine and of passages in his other works does not help us to interpret his claim.[1] It is possible that he believes that verbs can be rendered tenseless without a change of content or without a change of reference. Consequently, it is possible that he believes, not the contradictory of (B) when it is given a rules of use interpretation, but the contradictory of (B) when it is given a content or reference interpretation.

Since Goodman and Quine have not made their position clear, it is impossible to tell whether Braude has refuted it. If they claim that a rules of use version of the contradictory of (B) is true, then Braude has refuted their way of distinguishing between tensed and tenseless sentences. But if they claim that a content or reference version of the contradictory of (B) is true, then Braude has not refuted their position. Goodman and Quine should have distinguished clearly between rules of use, content and reference. They should also have stated that their position is a content or reference version of the contradictory of (B), since either of these versions is more plausible than a rules of use version.

All of this is relevant to 'such topics as the status of temporal becoming, subjective versus objective time, thing-ontologies versus event-ontologies, and perspicuous languages' (p. 188). Both Goodman and Quine make claims about some of these topics on the basis of the translatability of tensed sentences into tenseless sentences. Other philosophers have tried to refute their translatability theses so as to be able to make opposite claims about these topics. What is interesting is that some, if not all, of the claims made about these topics on the basis of the translatability of tensed into tenseless sentences can be made on the basis of both a content and reference translatability. Thus, those philosophers making opposite claims must try to refute not only a rules of use translatability (as Braude has done), but also both of these more plausible kinds of translatability (as Braude has not done). To illustrate: some philosophers, including Goodman, have said that temporal becoming is not a feature of physical reality because tensed sentences, which appear to embody the idea of temporal becoming, are translatable into tenseless sentences, which do not embody the idea of temporal becoming. Other

[1] Quine also discusses tense in *Word and Object* (Cambridge, Mass., and New York: Technology Press of M.I.T. and John Wiley and Sons, 1960), pp. 170–3, 191–4; *Philosophy of Logic* (Englewood Cliffs, N. J.: Prentice-Hall, 1970), pp. 30–1, 77; 'Mr. Strawson on Logical Theory', *Mind*, LXII (October, 1953), pp. 440–3; and *Elementary Logic*, p. 92. As on p. 6 of his *Elementary Logic*, it is not clear in these places whether he adopts a rules of use, content or reference translatability.

philosophers have tried to refute this translatability thesis so as to be able to maintain that temporal becoming is a feature of physical reality. These other philosophers must refute not only a rules of use translatability but also a content and reference translatability, since each of these entails that temporal becoming is not a feature of physical reality. For instance, if tensed sentences refer to the same states of affairs that tenseless sentences refer to, then those sentences (tensed ones) that appear to refer to temporal becoming do not actually do so, since tenseless sentences do not refer to temporal becoming at all.

The circumstance here seems to be analogous to attempted refutations of the materialist identity thesis. To refute that thesis, one cannot show simply that mental words and physical words have different uses. One must show that mental words and physical words have different referents. Doing the latter, however, is harder than doing the former. How often have we heard J. J. C. Smart and other identity theorists say, 'But you have only shown that mental words have different uses from physical words, and not that they have different referents'?

Possibly, those who make claims about the status of temporal becoming, such as Goodman, and about perspicuous languages, such as Quine, can reply to Braude, 'But you have only shown that tensed and tenseless sentences have different uses, not that they have different contents or referents'.

St. John Fisher College

LOOK, NO EYES

By CHRISTOPHER NEW

IN the exposition and defence of his general view that not all psycho-physical relationships can be merely contingent, Professor Sydney Shoemaker considers the particular case of seeing (*Self-Knowledge and Self-Identity*, pp. 174–9). Conceding that it may ordinarily be analytic that a person sees with his eyes, he holds that nevertheless 'there could be a form of perception which, whether or not the current use of the word "see" would permit us to call it seeing, is exactly like seeing except that it does not involve the use of the eyes' (p.174). I shall call this form of perception non-ocular vision. Shoemaker says of non-ocular vision that it is 'senseless' to suppose that it 'does not involve the perceiver's having a body *some* part of which plays the role that in ordinary vision is played by the eyes' (pp. 174–5; Shoemaker's italics). And the role he has in mind is that of providing a 'point of view'. Seeing, he argues, requires that there should be some point on the perceiver's body which is the point of view from which he sees. 'In normal vision, the point of view from which a person sees is the place where his eyes are' (p. 177); in non-ocular vision, there would have to be some other point on his body which was the point of view from which he saw. What is 'senseless', then, is to suppose that a person might non-ocularly see from a point of view which was not a point on his body.

Shoemaker tries to establish this conclusion by means of an argument which may be reconstructed as follows (see pp. 175–7).

> (i) A man might conceivably be able to make true statements about what is visible in places inaccessible to his eyesight. But while we might call such an ability clairvoyance, more would be required before we could call it (non-ocular) *seeing*.
>
> (ii) 'What is required is that he should speak and behave in such a way as to enable us to pick out some point on his body as the point of view *from* which he sees.' This implies that (a) There must be some point on his body 'such that he is normally able to give true descriptions of events occurring "in front of" that point, but is not able to do so if there is an opaque object between that point and the place at which the event occurs'; and (b) He must be 'able to make "egocentric" spatial statements in which words like "this", "here", "there", "near", "far", "left", "right", etc. are used to identify or locate objects or events in relation to the speaker; the truth or falsity of such statements depending in part on the way in which the things referred to are related to a certain point on the

speaker's body which is the "point of view" from which he sees'.

Hence

(iii) if a person claims to see something that is going on in regions inaccessible to his eyesight, his claim must be rejected (presumably as senseless) unless the conditions specified in (ii) (a) and (ii) (b) are satisfied.

Now Shoemaker makes no attempt to defend step (ii) in this argument; he merely asserts it bluntly—presumably because he thinks it is obvious. Yet it seems clear, upon a little reflection, that step (ii), with its conditions (a) and (b), is not only questionable, but false. Consider this example. A man in America claims that whenever he goes into the cellar of his house and concentrates very hard, he is able to "see" things going on in some other place which he has never visited and does not recognize. (He claims to "see" rather than see, because he is at the time utterly unable to tell whether he is having visions, hallucinations or is really seeing.) When asked to describe what he "sees", he is able to describe visual properties of (what seem to be) furniture, people and actions, all from a certain point of view. He says things like 'A heavy-set man seems to be now walking across the room, sitting down, coming towards my point of view,[1] passing out of my field of vision', etc. A team of investigators builds up a composite picture of the room from the man's descriptions and discovers that it fits exactly a room in the Kremlin which he has in fact never visited. They leave some of their investigators in the room with direct radio contact with the others, who observe the man and ask him questions. They find out that the man is apparently able to report only those objects or events in the room that are or occur in front of a certain point six inches above the floor and at the intersection of lines drawn three feet from the rear and left side wall. The man's range and field of apparent vision is determined by this point exactly as if he had actually been placed there with his eyes turned in a certain direction. When an opaque object is placed between this point and certain objects in the room, the man is unable to describe and reports he cannot "see" those objects, although he can "see" the obstructing opaque one. When an opaque object is placed behind the point, the man's apparent vision is unobstructed but he cannot describe and claims he cannot "see" where the opaque object is. When the room is

[1] It is not clear to me whether a man having such experiences would say something was moving towards his *point of view* or towards *him*. However, if he used the latter locution, he would not mean that the object was approaching his *body*. A helpful—though not exact—analogy is provided by the television screen. A man watching a live transmission sees the screen from the point where his eyes are, but he sees the events depicted on the screen from the point where the camera is. He may describe the plane he sees on the screen as coming towards *him*, although he is perfectly aware that it is not coming towards his *body*. (Strawson makes a similar point in *Individuals*, p. 91.)

darkened, the man finds it difficult or impossible to describe any visual properties and claims it is too dark to "see" clearly, or at all. After a time, the investigators decide that the man does non-ocularly see events and objects in the room from a certain point of view which can be precisely identified; and they go on to try to explain how he does so. Whether they are successful or not I shall not speculate, but whenever the man's claims are checked, they turn out to be correct—or at least as often correct as are the claims he makes to see things in the normal way by means of his eyes. And the man himself comes to speak of himself as seeing (not "seeing") what happens in the room.

While it is fanciful, there is nothing inconceivable about this example. The man's experiences and reports are exactly like those of a man seeing normally from the point of view identified by the investigators, except of course that no perceptions, apart from visual perceptions from just that point of view, and no perceptions of his own body, ever occur. In deciding to classify this as a case of non-ocular vision, the investigators are behaving perfectly rationally in the light of the facts. One of the facts, agreed, is that the man is aware of and discriminates the visual properties of things from a point of view which is not a point on his body. But that in no way impairs the intelligibility of the investigators' decision to classify his experiences as non-ocular vision. The man has apparently a field of vision and a point of view—and these, indeed, *are* logically required for seeing—but the point of view is simply not a point on his body.

Now if this fantasy is conceivable, Shoemaker must be wrong. And it is now possible to identify Shoemaker's fundamental error.[1] He has conflated two views, the first of which is true and the second false. It is true that in order for a person to be said to see, it is logically necessary that there should be some point which can be identified as the point of view from which he sees. It is, in particular, logically necessary that his perception of objects and events should be organized in relation to that point of view (though it is doubtful whether he must be able to make *statements* expressing this organization; for that would be to deny perception to non-statement-making beings such as animals and young children). That much *is* logically integral to the concept of seeing. But it is false that for a person to be said to see (non-ocularly), it is logically necessary that there should be a point *on his body* which is the point of view from which he sees. That one should see from a point of view on one's body—or, more exactly, from a point where one's body (or head or eyes) is, may well be necessary for the acquisition of normal (ocular)

[1] Not the *only* error. The denial in condition (ii) (a) of the logical possibility of seeing through opaque objects is neither defended nor defensible. It is easy to imagine a man's having "X-ray" vision and my example would not have been vitiated by giving the non-ocular seer that ability.

visual perception; but, once that has been acquired, there is no reason to suppose one could perceive *non-ocularly* only from a point where one's body or some part of one's (or any) body is.

It is interesting to compare the present example with Strawson's consideration of what we might call 'displaced vision' in *Individuals* (pp. 90–2). Strawson suggests it is logically possible for a 'subject of visual experience' to see from a point of view determined by the position of one body and the direction in which the head of another is turned, but only if the eyes of a third are open. This is a much more dramatic example than our present one and it may well be questioned whether a person (if such it is) so endowed, or afflicted, could intelligibly be said to *see* at all.[1] Much would doubtless depend upon whether Strawson intends to represent a temporary derangement of normal vision or the only mode of visual perception the subject has ever known—a point on which he is not explicit. If it is the latter interpretation that is meant, the prospects for intelligibility are dim—it is not even clear, for instance, whether such a 'subject of visual experience' would have a body to call his own; if not, the example may conflict with Strawson's own considered views on the concept of a person. However that may be, there is one feature of Strawson's suggestion which seems to me undoubtedly correct: it is not a necessary condition of seeing that one should see only from a point of view which is a point on one's body, or where (some part of) one's body is. This, indeed, seems to be the only feature which Strawson's example shares with the present one. So objections against the other features of his example would not stand against mine.

Another point of comparison is this. If Strawson's example is intelligible (as it might be, if taken to depict a possible derangement of normal vision), its subject would be most unfavourably placed to deal with, or even survive in, the world. But the circumstances of the man in our present example would be much more fortunate. His non-ocular vision would supplement his normal vision and be subject to his own control, rather than being involuntary and a derangement of his normal vision. Such a man might consider himself better, not worse, endowed, than the rest of us.

It will properly be asked whether Shoemaker's general thesis is refuted by the refutation of (this part of) his account of the psycho-physical relationships involved in seeing. The answer seems to be no. Shoemaker claims that not all psycho-physical relationships can be merely contingent and offers his account of non-ocular vision as an example of a non-contingent relationship in the case of visual perception. It has now been shown that this example is mistaken. It has not yet been shown, nor have I attempted to show, that his other examples are mistaken, or that the reasoning which leads him to his general thesis is

[1] See, for instance, B. A. O. Williams, *Problems of the Self*, pp. 124–5.

invalid. Nevertheless, there is a gap to fill: if a non-contingent psychophysical relationship must be involved in seeing, what is it? I believe the answer is to be found in the view that the notion of non-embodied perception is unintelligible. For if this is true, while it is conceivable that I should see from a point of view detached from my body, it is not conceivable that I should see without a body at all. Bodies are necessary to seeing, in other words, to provide, not a point of view from which one sees, but a perceiver who can have one. However, while I believe this answer is correct, I do not think there is space to defend it adequately here.

University of Hong Kong

LOGICAL SUBTRACTION AND THE ANALYSIS OF ACTION

By ROBERT A. JAEGER

I

WHAT is the meaning of Wittgenstein's question (*Philosophical Investigations*, I, § 621) 'What is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?'? A natural starting point is to define subtraction in terms of addition: the fact that P minus the fact that Q is that fact, x, such that x 'plus' the fact that Q 'equals' the fact that P. If we interpret 'plus' and 'equals' in terms of conjunction and logical equivalence we get this definition: the fact that P minus the fact that Q is the fact that R, where 'R' is that statement whose conjunction with 'Q' is logically equivalent to 'P'. Since it is more convenient to deal with statements than with facts we arrive at the following

D1. 'P' minus 'Q' (abbreviated '$P—Q$') = (def.) that statement whose conjunction with 'Q' is logically equivalent to 'P'.

The right side of D1 is a definite description, hence the question arises whether there is exactly one statement which has the relevant property. If there is at least one such statement then 'P' logically implies 'Q'; and if 'P' implies 'Q' then there is at least one such statement, for example, '$Q \supset P$'. But if 'P' implies 'Q' then both 'P' and '$(Q \vee S) \supset P$' (where 'S' is any statement whatever) also have the relevant property.

What is special about '$Q \supset P$' is that it is the weakest statement that has the relevant property: it is implied by all the others.

Those problems about existence and uniqueness which render D1 unacceptable suggest this revision

> D1a. If 'P' implies 'Q', then '$P-Q$' = (def.) the weakest statement whose conjunction with 'Q' is equivalent to 'P'.

So if 'P' implies 'Q' then '$P-Q$' is '$Q \supset P$'. This is the suggestion made in Hudson's critique[1] of my discussion[2] of Wittgenstein's question.

An important feature of the arm-raising example is that 'I raise my arm' *properly* implies (i.e., implies but is not implied by) 'My arm goes up'. But Hudson's theory of subtraction is quite general: '$P-P$' is well-formed and turns out to be '$P \supset P$'. Since subtracting a number from itself leaves zero, and subtracting a set from itself leaves the empty set, Hudson's theory is that necessary truths are the logical analogues of zero and the empty set. This may seem strange: one might find it more natural to think of necessary truths as the logical analogues of one, or infinity, and the universal set. Hudson will reply that necessary truths 'say nothing', but the answer to this is that such truths are true in *every* possible world. Everything depends on whether one is interested in intensions or in extensions.

When Hudson extends his theory to cover cases in which 'P' does not imply 'Q' two further problems arise. First, it is not universally true that 'P' equals '$P-Q$' plus 'Q', i.e., that 'P' is equivalent to '$(Q \supset P)$ & Q'. The latter statement always implies the former, but the former implies the latter just in case 'P' implies 'Q'. Hudson's definition of logical subtraction thus resembles the standard definition of set-theoretic subtraction: $A-B$ = (def.) $A \cap \bar{B}$. A is always included in $(A-B) \cup B$ (i.e., $A \cup B$), but the latter set is included in the former just in case B is included in A. In each of these cases one might argue that the operations in question do not deserve the name 'subtraction', since it is part of the meaning of 'subtract' that the minuend (whether a number, a set, or a statement) should be identical with the result of adding the subtrahend and the remainder. But Hudson will reply that what the objector takes to be part of the meaning of 'subtract' is merely an interesting property of arithmetical subtraction. The trouble with this particular dispute about the meaning of 'subtract' is that neither side is obviously right, and it is not clear what sort of evidence either side could invoke.

The second problem is that on Hudson's theory '$P-\sim P$' is well-formed and turns out to be '$\sim P \supset P$', i.e., 'P'. Now, it is obviously nonsensical to speak of subtracting the *fact* that $\sim P$ from the *fact* that P, since if it is a fact that P then there is no such thing as the fact that $\sim P$.

[1] James L. Hudson, 'Logical Subtraction', ANALYSIS, 35.4 (1975), 130–5.
[2] 'Action and Subtraction', *Philosophical Review*, LXXXII (1973), 320–9.

But Hudson is dealing with statements, not facts, and logical subtraction again turns out to resemble set-theoretic subtraction: observe that $A - \bar{A} = A \cap \bar{\bar{A}} = A \cap A = A$. Notice, however, that arithmetical subtraction once more turns out to be different: $x-(-x) = x + x$, and $x + x \neq x$ if $x \neq 0$. This draws attention to an important disanalogy between addition and conjunction: 'P & P' is always equivalent to 'P', but $x + x$ is not always equal to x. And one might argue that this disanalogy is so important that even if there is a logical operation whose relation to conjunction is the same as that of subtraction to addition, that operation does not deserve the name 'subtraction'.

It is impossible to refute Hudson's suggestion that the operation he defines deserves the name 'subtraction'. On the other hand, the relevant analogies and disanalogies are complicated enough to render groundless the idea that *of course* the operation in question is a genuine subtraction operation. The important thing is not to decide the terminological question, but to have a clear view of the considerations that make the decision so hard. As Wittgenstein remarks in another context (*Investigations*, I, § 79): 'Say what you choose, so long as it does not prevent you from seeing the facts. (And when you see them there is a good deal that you will not say.)'.

II

The topic of set-theoretic subtraction is more complicated than Hudson suggests. If we start with the idea that A minus B is that set, C, such that A is identical with C 'plus' B, then we can interpret 'plus' in terms of either union or intersection. Thus we get two rival definitions

D2. $A-B$ = (def.) that set whose union with B is identical with A.

D3. $A-B$ = (def.) that set whose intersection with B is identical with A.

Both D2 and D3 suffer from existence and uniqueness problems. As for D2, if at least one set has the relevant property then $B \subseteq A$; and if $B \subseteq A$ then at least one set—A, for example—has the relevant property. But if $B \subseteq A$ then $A \cap \bar{B}$ also has the relevant property. What is special about the latter set is that it is disjoint with B. As for D3, if at least one set has the relevant property then $A \subseteq B$; and if $A \subseteq B$ then at least one set—A, for example—has the relevant property. But if $A \subseteq B$ then $A \cup \bar{B}$ also has the relevant property. What is special about the latter set is that it is the most inclusive set that has the relevant property: it includes all the others.

These points suggest the following revised definitions

D2a. If $B \subseteq A$, then $A-B$ = (def.) that set which is disjoint with B and whose union with B is identical with A.

D3a. If $A \subseteq B$, then $A-B$ = (def.) the most inclusive set whose intersection with B is identical with A.

According to D2a, $A-B = A \cap \bar{B}$; according to D3a, $A-B = A \cup \bar{B}$.

The standard practice in set theory is to adopt D2a and then add that $A-B = A \cap \bar{B}$ even if $B \nsubseteq A$. It is to this definition of set-theoretic subtraction that Hudson compares his definition of logical subtraction. But '$Q \supset P$' is equivalent to '$P \vee \sim Q$', so the latter statement becomes the logical analogue of $A \cap \bar{B}$. Hence negation corresponds to complementation, disjunction to intersection, and '$P \vee \sim P$' to $A \cap \bar{A}$: necessary truths correspond to the empty set.

The alternative would be to adopt D3a and then add that $A-B = A \cup \bar{B}$ even if $A \nsubseteq B$. Were Hudson to exploit this unorthodox definition of set-theoretic subtraction '$P \vee \sim Q$' would become the logical analogue of $A \cup \bar{B}$. So negation, disjunction, and necessary truths would correspond to complementation, union, and the universal set. Thus the familiar link between disjunction and union would be preserved. Furthermore, comparing D3a with D1a would lead to the conclusion that one statement's implying another is the logical analogue of one set's being contained in another. This result is quite natural from an extensional point of view: if 'P' implies 'Q' then "the sense of 'P' contains the sense of 'Q' ", but the set of worlds in which 'P' is true is contained in the set of worlds in which 'Q' is true. The intensions and the extensions are oppositely related. So a weaker statement is true in a more inclusive set of worlds, and the weakest statements—those which say nothing—are true in every world.

III

'My arm goes up $\supset$ I raise my arm' would be unintelligible to someone who does not already understand 'I raise my arm', so if Wittgenstein's question is a request for an analysis of action then Hudson's answer is patently circular. Hudson might say that Wittgenstein's question is "purely logical" and is not a request for an analysis, but I think that that would be an oversimplification.

Subtraction and analysis are linked by the idea of a definition *per genus et differentiam*. Every case in which I raise my arm is a case in which my arm goes up, but not conversely. There is a species-genus relation here, and the question arises: what is the difference between those members of the genus that belong to the species and those that do not? This question presupposes that a thing belongs to the species just in case it belongs to the genus and possesses a certain *differentia*, which is to say that the species is the intersection of the genus with the set of things that possess the *differentia*. But since the question is a request for a definition the *differentia* must be free of circularity problems.

The idea of a definition *per genus et differentiam* lies at the centre of much philosophical writing about action. For example, Ryle suggests in *The Concept of Mind* that the Cartesian myth begins with such questions as 'What is the difference between intelligent and unintelligent behaviour?', to which the Cartesian replies that the critical difference is a difference in causation: intelligent behaviour is behaviour which is the effect of a mental cause. Ryle rejects this answer, 'the causal hypothesis', in favour of the view that intelligent behaviour is behaviour which is the exercise of certain dispositions. In *A Materialist Theory of the Mind* (London: Routledge and Kegan Paul, 1968) D. M. Armstrong rejects Ryle's answer in favour of a version of the causal hypothesis designed to escape Ryle's objections to the Cartesian version. Armstrong interprets Wittgenstein's question as the question 'what distinguishes' (p. 132) the rising of my arm, an example of 'physical behaviour' (p. 84), from my raising my arm, an example of 'behaviour proper' (p. 84); and he answers the question with a defence (pp. 131—144) of the suggestion that purposive activity is 'activity with a mental cause'.

Even if Armstrong escapes all Ryle's objections, the question remains whether Armstrong has provided an analysis of action. Armstrong's programme is to give 'accounts' (p. 84) of mental concepts in terms of concepts that do not 'presuppose mentality' (p. 84), thus rendering his accounts immune to circularity objections. But Armstrong's account of action is one half of a perfect circle whose other half is his account of perception (esp. pp. 250–255). The account of action exploits the concept of perception in that the latter concept is used in saying what the difference is between genuine purposive activity and the activity of a negative feedback mechanism. And the account of perception exploits the concept of action in that the latter concept is used in saying what the difference is between genuine discriminatory behaviour and a sequence of alterations in one's physical state which result from alterations in one's environment.

Armstrong deals with this circularity by emphasizing (pp. 252–255) the logical interdependence of the concepts of action and of perception. What he does not emphasize is that this interdependence destroys his programme. The holism of Chapter 11 is incompatible with the reduction of Chapter 6.

It seems to me, for familiar reasons, that Armstrong's holism is much more tenable than his reductionism. That is, it seems reasonable to suppose that the concept of action belongs to a circle of mental concepts each of which is analysable in terms of the others but none of which is analysable in terms of concepts outside the circle. What I wish to emphasize is that this interanimation of concepts implies that there is a sense in which the concept of action cannot be analysed at all, let alone reductionistically. If the object of the analysis game is to provide an *analysans* which

would be intelligible to someone who does not understand the *analysandum*, then the problem is that any adequate analysis of action will exploit other members of the mental circle, concepts that depend on the concept of action just as much as it depends on them. So one might say that the concept of action is weakly, but not strongly, analysable. If you command the concept of action and other concepts in its circle, then a philosopher can tell you how those concepts fit together; but if you do not know what action is, then a philosopher cannot tell you. The concept of a vixen, on the other hand, is analysable in both senses: it is easy to know what a fox is and what a female is without knowing what a vixen is.

So both Hudson and Armstrong wind up giving circular answers to Wittgenstein's question, the only difference being that Hudson's answer describes a smaller circle. If the concept of action is indeed not strongly analysable then the most that others can do is to describe larger and larger circles. Notice, however, that from this it does not follow that one can learn what action is only 'from one's own case', by a process of introspection and abstraction. Equally, the fact that one cannot learn anything from one's own case does not support the conclusion that it is possible to provide a non-circular analysis of the concept of action.

Yale University

TRACTATUS 2.0211—2.0212

By H. W. NOONAN

CONSIDER the following argument. If a proposition is a function of a proper name then if the proposition has a truth-value the name has reference (this is one of Frege's most fundamental theses). Accordingly, the assertion that the proposition is true or false (has a truth-value) rules out the non-existence of the things named by the name. But the assertion that a proposition is true or false can rule out no possibility; hence the non-existence of the thing named by the name must be *no* possibility, i.e., it must be an impossibility.

Let us consider the argument as applied to one of Frege's examples. If the proposition 'Kepler died in misery' is analysable as a function of the proper name 'Kepler', then if one says

It is true that Kepler died in misery

one rules out the non-existence of Kepler, and similarly if one says

It is false that Kepler died in misery.

So if one merely says

It is true or false that Kepler died in misery

one *still* rules out Kepler's non-existence. But the mere assertion that a proposition is true or false can rule out no possibility; hence the non-existence of Kepler must be *no* possibility, i.e., must be an impossibility—or else the Fregean analysis of the proposition 'Kepler died in misery' as a function of the proper name 'Kepler' must be mistaken.

It is this argument which is put forward by Wittgenstein in *Tractatus* 2.0211—2.0212:

> 2.0211. If the world had no substance, then whether a proposition had sense would depend on whether another proposition was true.
>
> 2.0212. In that case we could not sketch out a picture of the world (true or false).

(To see that this is Wittgenstein's argument one must recall (a) that for Wittgenstein all propositions are analysable as truth-functions of elementary propositions which are concatenations of names, (b) that for him, unlike Frege, a proposition's having sense is its being true or false, and (c) that he thinks of an elementary proposition as a function of the names contained in it (see 3.318).)

Wittgenstein thus concludes that there are necessary existents—things common to all possible worlds, and that these are the only things which can be named.

Whether this argument can be answered, and, if so, how, I shall not discuss here (briefly, however, I think the argument goes wrong in assuming a 'pure intermediary between the propositional *signs* and the facts'—see *Philosophical Investigations* I 94). Instead, I wish to point out that Frege himself is committed to something very like Wittgenstein's conclusion.

There is a fundamental disagreement between Wittgenstein and Frege over the question whether ordinary language is imperfect. Wittgenstein believed all his life that ordinary language was all right (see *Tractatus* 5.5563, *Blue Book* p. 28, and *Philosophical Investigations* I 98), although the conclusions he drew from this belief were different at different times. Frege, on the other hand, thought that ordinary language was imperfect, because it allowed the possibility of empty names, and so of propositions without truth-value. In a logically perfect language there would be no such possibility because it would never depend on the truth of a thought whether a proper name (with sense) had reference (see 'On Sense and Reference', in Geach and Black (edd.) *Philosophical Writings of Gottlob Frege* Basil Blackwell, 1970, p.70).

But how could this be guaranteed? Frege explains that it can be done by special stipulations—for example, divergent infinite series[1] can be stipulated to stand for the number 0, and, similarly, compound proper names constructed from the expression for a concept with the help of the singular definite article can be stipulated to stand for 0 if there is not a unique thing falling under the concept. Now, stipulations of this kind can guarantee that Frege's condition on proper names is fulfilled (i.e., that their having reference does not depend on the truth of any thought) only if (to take the second example) it is *necessary* that *either* just one thing falls under the concept whose expression occurs in the compound proper name *or* something is (i.e., is identical with) the number 0. For if this disjunctive proposition is not necessary then the stipulation's fixing a reference for the proper name will depend upon a certain possibility's not being the case, i.e., it will depend upon the truth of the thought which excludes that possibility. The form of the disjunctive proposition in question here is the following: 'Either just one thing is F or just one thing is G'. And, in general, a stipulation of the kind Frege suggests will guarantee a reference for a name only if some disjunctive proposition of this form is necessarily true. Now if there were no objects common to all possible worlds (as, presumably, 0 is)[2] this would entail that yet another condition would have to be fulfilled if stipulations of the Fregean kind were to be sure to guarantee reference, namely, that some disjunctive proposition of *this* form be necessarily true: 'Either

[1] This is what Frege *says*, but what he must intend to be speaking of is the *sign* for the limit of a divergent infinite series.

[2] It certainly is so for Frege.

there is some unique, contingently existent, object which is F, or there is some unique, contingently existent, object which is G'. But it seems most implausible that any such proposition could be necessarily true (it would entail, after all, that there *are* contingently existent objects, and this itself does not seem necessarily true), and I do not think that Frege would wish to argue to the contrary. But, if not, then the possibility of Fregean stipulations of reference guaranteeing reference independently of what thoughts are true depends upon the existence of necessarily existent objects. So if these are the only way reference can be guaranteed (and Frege never suggests another)[1] it follows that the possibility of a logically perfect language depends upon the existence of necessarily existent things (and a logically perfect language, remember, is just one none of whose sentences are either truth-valueless or possibly truth-valueless). It further follows that all the names of a logically perfect language have to be capable of naming necessarily existent things.

I hope it is evident how close this conclusion is to Wittgenstein's conclusion that the possibility of sketching out a picture of the world (true or false) depends upon the existence of necessarily existent objects which are the only things that can be named.

Of course, Wittgenstein's conclusion is stronger—his names *have to* name necessarily existent things; but both conclusions are far from the common sense view that there can be (because there are) perfectly good proper names which name contingently existing things and could not but do so. What accounts for the difference between the two conclusions is that Frege does not make it a requirement on the proper names in his logically perfect language that they be rigid designators (in Kripke's sense), but any rigid designators such a language contains will have to be names of necessary existents.

Trinity Hall, Cambridge

[1] Even in the *Grundgesetze*. The stipulations given there do not have the *ad hoc* appearance of the ones considered here (for example, a proper name constructed from an expression for a concept with the help of what Frege calls his 'substitute for the definite article' is stipulated to stand for the object falling under the concept expressed if there is just one, and otherwise to stand for the extension of that concept), but they do not differ from them in any way that makes my argument inapplicable to them. They are, of course, defective, because of Russell's Paradox, but this only strengthens my argument, if anything.

IS UNDEFEATED JUSTIFIED TRUE BELIEF KNOWLEDGE?

By JEFFREY OLEN

SINCE the appearance of Edmund Gettier's well known challenge to the traditional account of knowledge as justified true belief[1] questions concerning the sufficiency of the account have focused on cases in which *S*'s justification for believing that *p* is defective in a characteristic way, viz., cases in which a false belief plays a crucial role in *S*'s justification. Such cases have been characterized, alternately, as cases in which *S*'s justification for believing that *p* either depends essentially on a false belief or is defeated by a false statement.[2] And, it has been assumed, if a fourth condition to the traditional account were provided explicating what it is for a justification to depend essentially on a false belief or be defeated by a false statement, the so-called Gettier problem would be solved, and the traditional account of knowledge plus this fourth condition would constitute conjointly sufficient conditions for knowledge. I will argue, however, that such confidence is misplaced, i.e., that even undefeated justified true belief is not sufficient for knowledge, that there is a range of cases in which (i) *S* justifiably believes that *p*, (ii) it is true that *p*, (iii) all of *S*'s beliefs which support the belief that *p* are true, and yet (iv) *S* does not know that *p*.

In presenting the proposed counter-examples I will assume that justification is a matter of the relationship among beliefs. I will, that is, assume that *S* is justified in believing that *p* if and only if his belief that *p* stands in the proper relationship to his other beliefs. Accounts of justification construed along such lines have recently been proposed by Keith Lehrer and Gilbert Harman.[3] For Lehrer the belief that *p* stands in the proper relationship to *S*'s other beliefs if and only if, given these other beliefs, the proposition that *p* is believed to have a greater chance of being true than any statement which competes with it. For Harman the belief that *p* stands in the proper relationship to *S*'s other beliefs if and only if the proposition that *p* belongs to the conclusion of an inference which is warranted by *S*'s other beliefs, where an inference is so warranted if and only if it maximizes explanatory coherence while making the least change in *S*'s antecedent world view. For both, then, justification is a matter of coherence among beliefs. For Harman it is a matter of coherence among all *S*'s beliefs; for Lehrer it is a matter of

[1] Edmund Gettier, 'Is Justified True Belief Knowledge?', ANALYSIS, 23.6 (1963), 121–3.

[2] Keith Lehrer offers both characterizations in *Knowledge* (Oxford: Clarendon Press, 1974), p. 215. The characterization in terms of a justification being defeated by a false statement was first offered, by Lehrer and Thomas Paxson, Jr., in 'Knowledge: Undefeated Justified True Belief', *Journal of Philosophy*, LXVI (1969), 225–37.

[3] Lehrer, *Knowledge*, ch. 8; Harman, *Thought* (Princeton: Princeton University Press, 1973), ch. 10.

coherence among only those beliefs which *S* would hold were he a disinterested seeker of truth.

Before proposing my counter-examples I should note that they are directed, not only to Lehrer and Harman, but to any undefeated justified true belief theory in which justification is construed as coherence among beliefs. The proposed counter-examples are as follows:

(i) John is working on a crossword puzzle, and only one square remains to be filled. By filling in this last square he will complete two words. The square will contain the third letter of a three-letter word going down, the first two letters of which are *n* and *a*. He has encountered the clue in previous puzzles, and recalls that the answer is either 'nag' or 'nab', but cannot remember which. The square will also contain the first letter of a three-letter word going across, the second and third letter of which are both *e*. The clue is 'right turn'. He knows that the answer must be 'bee' or 'gee'. Both terms appear in larger expressions which seem to be relevant to his deliberations. 'Bee' appears in 'beeline', meaning a straight and direct course; 'gee' appears in 'ogee', meaning an S-shaped curve. On the basis of these larger expressions, given a choice between *b* and *g*, he infers that the latter must be correct. He infers correctly.

(ii) John is a contestant on a television quiz programme. The host asks him the following question: 'Cromwell when leading a battle charge used to exhort his men to battle with the following cry: "Trust in God and keep your (blank) dry!" What did Cromwell tell his men to keep dry?' John thinks about the question, running through all the things a man might do well to keep dry in battle. Knowing that Cromwell was not given to jocularity in such circumstances, he quickly rejects such candidates as 'feet'. Knowing also that Cromwell's men would have used powder as ammunition, and knowing in addition the importance of keeping powder dry, he infers that 'powder' must be the correct answer. And once again, he infers correctly.

(iii) The year is 1995. John is asked who won the National Baseball League's Most Valuable Player Award in 1975. He cannot remember who won the award that year, but he recalls that the Phillies won the pennant that year, and that Greg Luzinski led the league in home runs and runs batted in, while ending the season with a batting average above .300. (This is all wishful thinking, of course, but, for purposes of the example, true.) Knowing that the awarders of the MVP like to give it to a member of the pennant winning team if there is a suitable candidate, and knowing also that Luzinski was the best hitter in the league in 1975, John infers that he must have won the award. Once again he infers correctly.

In all three cases John's conclusions meet the conditions of knowledge set by Lehrer and Harman. Thus, on their accounts, we must say

that John knows the answers to the above questions. But John does not know the answers. We do not ascribe knowledge in such cases; we say, rather, that the person made a good guess, or an educated guess, or that he inferred correctly, or something of the sort. If John had learned the answers previously from a reliable source and then reproduced them from memory, then he would be said to know the answers. As it is, however, he does not know them, no matter how certain of them he may be, and he will not know them until he consults a reliable source—a dictionary, for instance, or a history book or the back issues of the sports pages, depending on the case. His warranted inferences to true conclusions do not, in these cases, yield knowledge. They yield the most reasonable hypotheses, but these hypotheses stand in the need of confirmation. They do not yield knowledge because, even though John is justified in believing them, his justification is not sufficient for the given situation. In each of the three examples there is something more that John can do in order to confirm his hypothesis, and it is the sort of thing we normally expect people to do in such cases.

Thus accounts such as Harman's and Lehrer's fail to make an important distinction, the distinction between an undefeated justified true belief that constitutes knowledge and an undefeated justified true belief that constitutes the most reasonable hypothesis. Furthermore, our conclusion seems to remain unaffected if we expand our scope to include another type of account in which justification is construed as a relationship among beliefs, viz., a foundationalist one. Given the above sort of cases, it is difficult to see how the necessary distinction can be made by an account that considers a belief justified if it is warranted by a person's other beliefs, whether such justification is a matter of coherence or of a line of support terminating in basic knowledge. What is needed, apparently, is an additional condition to undefeated justified true belief. This condition must require that *S*'s justification be sufficient to yield knowledge in the given situation. It cannot, however, be formulated in these terms, since such a formulation is obviously trivial.[1] It must be formulated in such a way as to allow an independent determination of the sufficiency of *S*'s justification in the given situation. And, considering the vast amount of contextual considerations which bear on such a determination, the prognosis for such an account is not promising.

Temple University

[1] For such an (admittedly) trivial account, see James Cargile, 'On Near Knowledge', ANALYSIS, 31.5 (1971), p. 146.

WHY THE UTILITARIANS SHOT PRESIDENT KENNEDY*

By DON LOCKE

IT now appears that sometime in the early 1960s, when other moral philosophers were agonizing over the logic of moral judgements and the meaning of 'good', a group of Utilitarians was meeting in the United States to discuss the practical applications of their theories and the good they might do in the world. Among their principal concerns was the evident increase in inner-city violent crime, which seemed to them to threaten the very roots of civilized life, and they placed the cause fairly and squarely in the traditional American policy of ease of access to firearms. Yet they knew that public pronouncements and political pressures from a handful of academic moralists would count for nothing against the powerful gun-lobby in Congress and the ingrained attitudes of millions of their countrymen. The only hope of change lay in shocking a nation to its senses by demonstrating how easy it was to purchase a gun on any American main street, or through the mails, and shoot anyone that might take your fancy.

Accordingly they decided on the course of action which seemed most likely to produce the desired benefit at least cost, one which might appear bad on the surface, but which would ultimately prove more productive of good than evil. They decided to shoot just one individual, so as to minimize the inevitable suffering and misery, but the one individual whose prominence and popularity—personal if not political—would be most likely to horrify the country and lead in a mood of national mourning and self-criticism to the imposition of stringent gun controls.

Even so the conspirators had their qualms. It was not simply that their Utilitarianism was permitting, indeed requiring, an action which on any non-Utilitarian ethic must seem horribly and unquestionably wrong: as Utilitarians they had long been inured to that familiar possibility. But Utilitarianism was requiring an action *because* it was wrong, inasmuch as its eventual rightness depended precisely on its being universally condemned, even among Utilitarians. But they comforted themselves with the thought that what was necessary was not that their action be wrong, but only that it seem wrong. Evidently a Utilitarian must be prepared to make the right seem wrong, and also, presumably, to make the wrong seem right: unlike Justice, Utility need not be seen to be done.

And so the deed was done, for the best of Utilitarian reasons and for the best of Utilitarian results. But unhappily things went spectacularly wrong. Not only did the desired results fail to accrue, but it is quite

* I uncovered this plot while reading *Utilitarianism: For and Against* by J. J. C. Smart and Bernard Williams (Cambridge, 1973), and especially Section 6 of Williams' paper. Further evidence may be found in 'It Makes no Difference Whether or Not I do it' by Jonathan Glover, *Proceedings of the Aristotelian Society, Supplementary Volume* XLIX, 1975.

possible that this incident was largely responsible for the subsequent cult of political assassination. Certainly the outcome was that violent death became even more an accepted part of the American city scene. But although the conspirators' action proved disastrously wrong in objective terms, it is difficult for a Utilitarian to deny its subjective rightness. The plotters may regret the way things worked out, but it would be irrational of them to feel any remorse or guilt—or whatever the Utilitarian equivalents may be. They miscalculated the strength of the gun-lobby, the depth of the average American's resistance to any interference with his right to carry firearms; but if any event chosen so as to involve least harm in itself seemed likely to achieve the desired result then that was it, and we are able to question it only with benefit of hindsight. They took a calculated risk, such as any agent must do when estimating the consequences of his actions, and if they failed it was not through any fault of their own: the probable good from their action far outweighed the probable harm. Whether other Utilitarians should praise or blame them must depend, of course, on the utility of such praise or blame, and no doubt they were worthy of blame, *pour décourager les autres,* even if they had been successful. But no Utilitarian will find moral fault in their acting as they did in the circumstances, and with the wholly reasonable expectations, that they did. They were eminently blameworthy, but by no means bad.

In another respect, moreover, the Utilitarians have been much less in the wrong. It was, of course, morally crucial to their plan that no one discover why or by whom the deed was done. The latter would only increase the resulting harm, when the conspirators paid the appropriate Utilitarian penalty for their good deed; and the former might bring Utilitarianism itself into such public disrepute as to hinder future Utilitarian policies and practices. For a moment, indeed, it seemed as if the plot might fail here too, and be uncovered, with its Utilitarian motives and justification laid bare for all to see and reflect upon. But happily for the prevention of evil, it proved possible to divert attention to a solitary individual, again minimizing the inevitable harm, and to avoid the risks of exposure inherent in a public trial by due legal process. In fact it proved easy enough to convince the authorities appointed to enquire into these incidents, and to set public doubts at rest, of the Utilitarian importance, which is to say the moral value, of protecting the Utilitarian secret; and though suspicions have remained rampant, no one has so far come near to tracing the plot back to its real source. The original action may have proved an unqualified failure, but all men of good will would be able to rejoice that to date the cover-up has been an unmitigated moral success.

But now the jig is up, the word is out, and the plot has failed here too. Indeed if people are led to question an ethic which, but for a bit of bad

luck, would have justified the murders of two innocent men and all the subsequent political intrigue and public deception, then I fear this note is guilty, not only of bad taste, but of moral subversion. But I have two defences.

One is that the fault cannot wholly be mine, if the same evil would have followed without my intervention. In fact the Utilitarians themselves are equally to blame. For generations they have been publicizing a morality which insists that the value of a deed depends not on its reasons but on its results. So even if the assassination had been the work of the unfortunate who, for good Utilitarian reasons, was made to take the blame, his action could still have been justified, would still have been right in Utilitarian terms, if only it had had the desirable result of a drop in violent crime via firearm controls. His acting thus, supposing he had acted thus, would then have been more productive of good than his not acting thus, and worthy of honourable mention in the annals of great twentieth-century moral achievements (not that such annals should ever be made available to public inspection). This conclusion was there to be drawn, even by those with no inkling of the Utilitarian plot; sooner or later the truth—about Utilitarianism, if not about Dallas—would have emerged. There are, then, good Utilitarian reasons why a Utilitarian should not preach his Utilitarianism, except to those so secure in their Utilitarianism that no consequence thereof will lead them to question it. But driven by a natural human desire to make converts, rather than doing good in the world, the Utilitarians have not been content to follow this precept, and they must bear some of the responsibility.

Of course it is possible that the conspirators were mistaken in thinking that the likelihood of people's performing good Utilitarian acts would be affected by their acceptance or rejection of Utilitarian principles; and equally it is possible that no Utilitarian harm will be done by publishing the consequences of Utilitarianism. In that case the preaching of Utilitarianism will not be harmful, merely idle, though there will still be no good Utilitarian reason for publicizing Utilitarianism. That at any rate is my hope and my second defence: that however much the morally unsophisticated might be appalled by the deeds that Utilitarianism can be used to justify, and the deeds that might be done in its name, a modicum of ignorance, deception or human perversity may yet suffice to ensure that everybody acts for the Utilitarian best in the best of Utilitarian worlds.

University of Warwick

REPLY TO GRIM

By LARRY WRIGHT

I DO not think Mr. Grim's contention that a clear philosophical outline of functions is still a long way off gains much support from the counter-examples he brings against my analysis (ANALYSIS 35. 2, 1974). I have never thought of the counter-example as a purely destructive device, and given the complexity of the concepts philosophers want to analyse, it would seem to be a necessary tool for analytic progress. Such progress is very often by successive approximation, trial and error, making *use* of counter-examples at every step. And whether a given step is a misstep, or an important insight, depends in part on what specific difficulty the counter-examples point to, how they can best be accommodated. Grim's attack on my analysis of function provides a useful illustration of this.

I had claimed (*Philosophical Review*, April, 1973) that having a function consisted in having a certain sort of causal history, which I called a 'consequence-etiology' and formulated explicitly as follows.

> The function of X is Z if (a) X is there because it does Z and (b) Z is a consequence of X's being there.

My original intent was to display functional ascription-explanations as explanations of how the thing with the function 'came to be there'. Accordingly, the 'is there because . . .' of (a) would have to be read something like 'got there because . . .' This would accommodate the standard cases of natural and conscious functions. In the former case, propitious first mutations would be disqualified for a few generations, until the advantage of the mutation could claim some responsibility for its surviving, but thereafter would fit the formula on its original reading. The feature would then be there (in the particular organism) at least in part because of what it did. Standard cases of conscious functions fit in a straightforward manner.

However, there are some cases of conscious functions which do not comport with the formula on its 'got there because . . .' reading. In the article Grim refers to I consider the following case.

> . . . if an earthquake shifted the rollers of a transistor production-line conveyor belt, causing the belt to ripple in just such a way that defective transistors would not pass over the ripple, while good transistors would, we could say that the ripple was *functioning as* a quality control sorter. But it would be incorrect to say that the ripple *had* the function of quality control sorting. It does not *have* a function at all. It is there only by accident. Sorting can, however, *become* its function if its sorting ability ever becomes a reason for preserving the ripple: if, for example, the company decides against repairing the conveyor belt *for that reason* (p. 166).

To accommodate cases like this, I relaxed the formulation slightly and encouraged a "maintenance" reading of 'is there because'. The ripple did not *get* there because of what it does, but it has *stayed* there because of what it does. This seemed good enough at the time.

Unfortunately, it is just this relaxation which allows Mr. Grim's counter-examples. This defect was first brought to my attention by Mr. N. McKenzie of York University when the analysis was originally published. He used the following perspicuous example.

> Jones is thrown in jail overnight for a minor offence. He is so constituted that being in jail makes him scream obscenities. Unfortunately, there is a law that an enforcement officer may confine anyone who continually screams obscenities. So Jones is kept in jail indefinitely. Now: (a) Jones is there (in jail) because he screams obscenities. (b) His screaming obscenities is a consequence of his being in jail. But Jones' function is not to scream obscenities.

What examples like this show, I think, is that it was a mistake to attempt to accommodate cases like the conveyor-belt ripple as paradigms. The paradigm conception *is* 'got there because of what it does', and the ripple type cases are best viewed as derivative. If this move is made (that is, the "maintenance" reading discouraged), the McKenzie-Grim counter-examples are avoided. But perhaps more importantly, we also gain some insight.

For to defend ripple-type cases, even as derivatives, involves taking advantage of the subtle, subjunctive causal possibilities inherent in conscious intent and decision. The thing did not get there because of what it does, but, ceteris paribus, it *might well* have, had it not done so accidentally.[1] It has a subjunctive consequence-etiology. But this notion has no close parallels in natural (evolutionary) cases. So a second aspect of the insight to be gained here is that conscious and natural functions split apart somewhat on the borderline cases. This is just as we should expect, and it is helpful to see it in some detail.

University of California, Riverside

[1] A complete account will be available from The University of California Press this winter, entitled *Teleological Explanation*.

WHAT'S WRONG WITH DISCRIMINATION?

By PAUL WOODRUFF

THE trouble with compensatory discrimination in favour of a group, say its critics, is that it is discrimination. It is morally wrong because it accords differential treatment on morally irrelevant grounds—it distributes reparations on the basis of group membership, whereas if reparations are to be made, they should be made to the individuals who have been wronged. (William A. Nunn, ANALYSIS 34. 5, pp. 151—154, and Alan H. Goldman, ANALYSIS 35. 5, pp. 169—170.) The defenders of compensatory discrimination for groups tend to deny that its criteria are morally irrelevant. A person's membership in a group, they say, may become morally relevant when there has been systematic discrimination against that group. (Paul W. Taylor, ANALYSIS 33. 6, pp. 177—182.)

No one, it seems, stops to consider the relevance of moral relevance to the issue. The important question is why discrimination is wrong when it is wrong. When we know the answer to that we shall know who it is that is wronged by it.

For brevity's sake, and because it is the hardest case to make out, I shall consider only discrimination in hiring by private employers. The problem is that it does not always seem to be wrong. Suppose a private banker hires his cousin in preference to a better qualified applicant from another family. Hiring bankers this way is discriminatory, for it uses a criterion (kinship) that is irrelevant to banking. Is it morally wrong?

We need to know more about the case. Suppose that our banker's bank is the only bank in the country, that it is entirely a family affair and that it dominates the modern economy of the country. Then, whenever our banker hires a cousin, he helps to exclude those outside his family from power and to raise over them a class with birthright privileges. He participates in a pattern of discrimination that wrongs *everyone* who is not a member of his family by making their parentage a reason for failure. This is insulting to every member of the excluded class, and damages their self-respect and the respect in which they are held by others.

On the other hand, if our banker's bank was a small business in a land of many and various banks, which were not in general ruled by a privileged class, then what moral wrong would he commit in hiring his cousin? If the cousin is idle, it is foolish to hire him. If an applicant is disappointed, it is sad not to hire him. But I am not aware that it is generally wrong to disappoint people.

The disappointed applicant may have valid complaints against the banker that have nothing to do with discrimination. For example, in advertising the job, our banker may have promised to give it to the best

applicant. Or using customary hiring practices like interviews may oblige him to hire the best applicant. But he can discriminate without violating obligations of that sort. Also, it may turn out that no one hires the applicant. Whatever the cause of this, and whether or not discrimination was involved, his unemployment may be a social injustice. If so, the banker does wrong by contributing to that injustice. But discrimination and the fate to which it may bring its victims are wrong in different ways. For example, it does not matter how people are made slaves; their slavery is an injustice whether they came to it by stupidity, conquest, discrimination, or an impartial lottery. Slavery is an injustice to the slaves. But enslaving people *for their race* would add a wrong to slavery, and not to slaves only, but to every member of the insulted race.

Notice that the moral relevance of the criteria used is not decisive. Kinship would be morally relevant to our banker's decision if he himself had benefited from the hiring of cousins. But what really matters is whether or not it is right to hire cousins.

I suggest that an act of discrimination is wrong when it is wrong not simply because it is discriminatory, but because it is part of a pattern of discrimination that is wrong. A pattern of discrimination is wrong when it makes membership in a group burdensome by unfairly reducing the respect in which the group is held. It may accomplish this, for example, by making group membership a *prima facie* reason for failure. One act of discrimination cannot do that. If an applicant fails at one bank because of his race, and at other banks for other reasons, his race is not the reason for his unemployment, and his failure is not an insult to his race. Discriminating, like walking on the grass, is to be judged with reference to how much of it is being done. Walking on the grass is harmful only if enough people are in the habit of doing it to ruin the grass. So it is with walking over the feelings of a group.

My account of discrimination yields two conclusions for the debate on compensatory discrimination.

(1) The objection that compensatory discrimination is wrong because it is discriminatory may be dismissed. Compensatory discrimination is not part of a pattern that is wrong. Since it is compensatory it is limited to a scale which will not reduce *unfairly* the respect in which the compensating group is held. Though it may cause certain people to be unemployed, it does not increase unemployment, which is an evil however it is distributed.

I do not consider the objection that compensatory discrimination violates the principle that the best people deserve the best jobs (the meritocratic principle). The merits of that principle are much in question, and raise too large an issue for this brief note. In any case, the chief complaints against discrimination are based not on that principle but on

egalitarian beliefs. I have tried to make the burden of those complaints precise.

(2) It is now clear that compensation for discrimination is owed to all members of the relevant group in virtue of their membership in that group. An individual disappointed by discrimination has no special claim to redress on that score. He has been wronged, not because he has been disappointed, but because he is a member of a group that has been wronged. Every member of the group finds membership in the group unfairly burdensome as a result of the pattern of discrimination; every member has an equal claim to compensation. Conversely, every member is benefited equally by an act that tends to break the pattern. Our banker would do well by the non-cousins if he hires even a privileged non-cousin, one who has not yet been disappointed, if in doing so he increases the accessibility of bank jobs to non-cousins.

Of course in the history of any pattern of discrimination there are usually at least two wrongs to be righted. Thus if people have been enslaved on the basis of race, then (i) every member of the race deserves redress for discrimination, and (ii) every slave deserves redress for enslavement. Slavery, poverty, and the like are borne by individuals; discrimination is borne by groups.

But what precisely is a group for our purposes? Consider the group of those who are either black women or hold degrees in law from Harvard. Few of its members in proportion to its number enjoy jobs of importance. Yet no white Harvard lawyer is wronged by the snubbing of a black woman, and it would be no consolation to black women to see a Harvard lawyer advanced to a good job, if he is white. We need to restrict what counts as a group for our purposes. As consistent egalitarians, we must avoid certain moral and natural criteria for grouphood, on pain of undermining our objection to discrimination. The relevant group in each case is the group that has been insulted. What that is is a question for social scientists.

The University of Texas at Austin

NOTES

SUBSCRIPTIONS. The subscription to ANALYSIS for institutions is £5.00 (inland), £6.00 (overseas), $15.00 (U.S.A. and Canada); for individuals £4.00 (inland), £4.80 (overseas), $12.00 (U.S.A. and Canada). Each volume comprises four numbers, three of 48 pages and one of 64 pages, appearing within the academic year—in October, January, March and June. Orders should be sent to Basil Blackwell, 108 Cowley Road, Oxford OX4 1JF, or placed with any bookseller.

CONTRIBUTIONS. Articles submitted for publication should be addressed to Dr. C. J. F. Williams, 1 Fossefield Road, Midsomer Norton Bath, BA3; 4AS. Contributors are asked to note the following requirements.

Articles should **normally** not exceed 3,000 words in length. **Occasionally,** however, longer contributions can be accepted;

They must be type written in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

It is regretted that owing to increased postal charges it is no longer possible to return typescripts unless the following instructions are followed:

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgement is required, a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope **and international reply coupons** of the requisite value, whether for air or surface mail.

Galley proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts will be retained by the Editor on the assumption that authors have kept their own copies.

ISSN 0003-2638

PRINTED IN GREAT BRITAIN BY BURGESS & SON (ABINGDON) LTD., ABINGDON, OXFORDSHIRE

Vol. 36 No. 4 (New Series No. 172) June 1976

ANALYSIS

Edited by
C. J. F. Williams

CONTENTS

BASIL BLACKWELL · ALFRED STREET · OXFORD

Price £1.20

ANALYSIS 36:4 JUNE 1976

ON FRUSTRATION OF THE MAJORITY BY FULFILMENT OF THE MAJORITY'S WILL

G. Ξ. M. ANSCOMBE

WHERE matters are decided, yes or no, by a majority vote, the decision on each matter is the will of the majority. But it is also possible that the majority should vote in the minority on a majority of the questions. This may be shewn in a table.

	A	B	C	D	E	F	G	H	I	J	K
1	0	1	1	1	1	0	0	1	1	0	0
2	0	1	0	1	1	0	0	1	0	1	1
3	0	1	0	0	1	0	0	1	1	1	1
4	0	1	0	0	0	1	1	0	1	1	1
5	0	0	0	0	1	1	1	0	1	1	1
6	0	0	0	0	0	1	1	1	1	1	1
7	0	0	0	0	0	1	1	1	1	1	1
8	1	0	0	0	0	0	1	1	1	1	1
9	1	0	1	0	0	0	0	1	1	1	1
10	1	0	1	1	0	0	0	1	1	0	1
11	1	0	1	1	0	0	0	1	0	1	1

Here we have eleven voters, A–K, voting on eleven questions. Seven of them, A–G, vote in the minority in a majority of the decisions: A–F in seven out of the eleven cases, G in six. The majority is always 6–5. These figures can of course be varied.

If we imagine an ideal democracy with a whole population voting directly on all questions, there will obviously be room for much variation in results over a long period, all of which however conform to the description: the majority votes in the minority in a majority of cases.

This fact, I thought when I stumbled on it, must be familiar to voting experts. But I have not found it remarked upon. It sometimes startles people, eliciting the reaction: 'But doesn't this make nonsense of democracy?'

In the West, and perhaps in the whole world where Western forms of education prevail, men are brought up in a conviction of the unique fairness of democracy. It is even conceived to be as it were the sole legitimate form of government. 'It's not democratic' is a condemnation. Pope Pius XII once spoke in a Christmas allocution of the right of democracies to defend themselves by whatever means they might think necessary. One could hardly have better proof of the pervasiveness of

the attitude, if of all people the Pope of Rome should speak as if democracies had some special entitlement. Outside governments, we notice that methods of proceeding and alterations in arrangements are often proposed on no other ground than that they will be "more democratic". I know a university, for example, where on these grounds representation in the authoritative assembly was made proportional to the numbers in the different departments. Large departments proceeded to forbid expansion to or even to reduce small ones, increasing themselves as every opportunity offered.

In the Liberal tradition of the West, democracy has certainly been favoured, and belief in it instilled as a fundamental dogma. But there are other strands belonging to that tradition. One of them is a concern for minorities. Upholders of this, it seems, were conscious that a majority may be a cruel and oppressive bully to a minority. In consequence "protection of minorities" or "of the rights of minorities" also had a certain sacred flavour. This it perhaps sometimes retains more in the United States than in England, where a common denigratory phrase nowadays is 'merely sectional interests'. However this may be, I notice in my own country at least a hardening of sentiment exclusively in favour of democracy, which is sometimes also called 'majority rule'. Such is the sentiment in favour of this, that it is not *astonishing* to hear of an apparently reasonable and well-disposed man who says: someone who is not prepared to accept a majority decision, or a law enacted by democratic processes, ought to leave the society. And this, without any limitation's being put on what the matter of the law or decision may be. We see that from such a mind the Liberal tradition of concern for the "rights of minorities" has been nearly expunged.

Let us consider the *rationale* of the idea that decision according to the will of the majority is superior to any other. That a particular decision according to the will of a majority may be inferior, in the sense of stupider, is evident. I take it that a decision is stupid if its implementation is undesirable from the point of view of the very people who wanted it. They would have it: but when they get it, it is evident that they would rather not have it; if they had only realized what it would be like for their decision to be implemented, it would not have been made. There was another decision available, perhaps: if the power of making the decision had not been taken away from some particular person and vested in the majority. And it is now recognized that *that* decision would have been better.

The superiority of the majority decision cannot reside in its necessarily superior wisdom. A superiority, even when the decision is stupid, is rather thought to be in this: participation in the decision-making is itself valuable; or, perhaps one should say, is itself *a value*. Such a value is mystical of course. I don't mean a criticism in saying so. But

consider a man who always or most often finds himself voting in the minority, and who judges that most of the decisions taken are extremely harmful. Why may he nevertheless extol this method of arriving at decisions? He may think that alternatives available in the situation would lead to even worse ones, but that is not to put a value on participation as such. He may feel some satisfaction with the method in that there *was* always a contrary voice, and that his made part of it—but why? Honour satisfied? That *is* mystical. A nucleus of possible opposition and difference of direction? There may be none such: the minorities in which he votes may be inconstant sets of people. Does he simply have a "we" feeling about the decisions, bad as he thinks they are? '*We* decide for ourselves'?—even though *his* vote was always contrary to the decision? There is indeed such a sentiment, but, once again, it is mystical. He has reason to feel as oppressed by the authority of the majority as he might by the authority of an autocrat or superior body, who "hand down" decisions that he has to accept.

Now there is a quite different belief to yield the superiority of majority decision except, indeed, where it is stupid in the sense I have explained. Let us suppose a set of people who have to determine, or for whom it must be determined, what they shall do or what shall happen to them collectively. Suppose that they are all going to go somewhere in a vehicle; the question is, where? Some would like to go to one place, some to another. They can't go to both. But they must all go together where they go. Doesn't it then seem reasonable that they should go where the majority of them want to go? All cannot be satisfied; in this way, perhaps, most of them will be. Complications arise, of course, where the choices are multiple; but let me assume just a simple choice, made by everyone, between a pair of alternatives. Then the majority, if there is one, will be absolute. Now it does seem most reasonable that, where all cannot have their way, the greater number should not be frustrated.

Note that *this* defence of majority decision in conveniently simple cases does not assume that there is actually a value in people's sharing the job of making the decision. For the argument only concerns people's getting what they want.

The reasonableness is meant to consist in a sort of fairness. If the desires of the majority are not to prevail, then a few perhaps get what they want at the expense of a greater number who do not. This seems to put the desires of those few in a privileged position; which is intrinsically unfair. When the majority get what they want, they do so at the expense of the minority's frustration, but at least nobody is privileged; for the upshot results from making the desire of each count equally. *Ex hypothesi* no upshot could satisfy all equally; a method which derives the upshot from an equal weight attached to each person's desire must

achieve the satisfaction of the desires of the greatest possible number and so be both the best and fairest.

This belief about the best and fairest upshot is independent of the method adopted to secure that upshot. If the method adopted can be that of majority vote on simple alternatives this, it is arguable, is the best method because presumably people are themselves in the best position to know what they want. Even though they may sometimes make mistakes, that is worth putting up with, because they will surely, at least for the most part, vote for what they really do want. But the method of majority decision is here espoused, not because of any intrinsic value in people's sharing in making decisions, but on the assumption that it is the best available method for securing the upshot. The argument that this is the best and fairest upshot does not depend on this being the method for securing it; but, stupidity apart, this method would secure it more surely than any other.

Note that that way of looking at the merits of majority decision starts off with an assumption: Namely, that a decision has to be made for people *as a collection*; a decision has to be made which determines what everyone does or has in some matter. This was why I supposed the case where it was given that a party of people were in for going somewhere in a vehicle. We should distinguish between this set of people and a set of people in a house who happen to have a practice of determining daily by majority decision what all of them are going to do that day. Here one might say: 'Why not let each choose his own activity?' and the argument I gave purporting to show that the will of the majority should prevail will have no force.

At least, it will have no force until we take into account that some people's desired activities include making other people act in various ways. The philosopher that I referred to, who thought that one had no right to remain in a society if one was not prepared to accept its majority decisions, did not stipulate that the decisions should concern only actions and fates where a decision for a whole collection of people is necessary. This is perhaps realistic for the following reason: some people desire to control others, to dictate to them, as also to get them (in various ways) to do things they would not otherwise do. Now if our initial assumption is: each person's desire shall have as much weight as anyone else's in calculating what to do—then these types of desire can't be excluded from consideration. As soon as there is a proposal to interfere with X's activity, someone presumably wants to do so. If interfering can come up for majority decision at all, and it is proposed, some want and some perhaps do not want a decision that that set of people collectively interfere with X. Even if those prevail who do not want to interfere with X and do not want it to be a collective decision to interfere with X, it has come about that there necessarily is a collective

decision *whether* to interfere with X.[1] What reason could be offered against the paramountcy of the will of the majority? The "given", that they are "in for" some decision (as it was a "given" that my set of people were going somewhere in their vehicle), has been contrived merely by the question's coming up for majority decision at all.

Contrast the situation of an autocrat; he is the source of all the decisions about collections of people. He is also so powerful that he can interfere with any arbitrary X if he likes. If he does so (or abstains) his action remains an action just relating to X; it is not transformed into a decision about people collectively.

I do not want to steer further out into these deep waters. For my quarry is something quite different. We had an argument purporting to show that where there are a lot of people collectively in for one of a pair of alternatives, the best and fairest decision was the alternative preferred by the majority. It seemed *best* because we are envisaging no criterion of what is good except that it is wanted, or of evil except frustration of desire; it seemed *fairest* because everyone's desire is given equal weight. The tradition of concern for minorities is nourished by far different considerations than these. What those may be is not at all my concern. I am impressed by the argument, even if I would like to draw from some other source some principle of limitation of its applicability.

But now—the table presented right at the beginning seems to damage this argument, which at first sight seems so impressive. For what do we have here? Each decision is made according to a majority vote. Each decision, then, (folly apart) would seem to be the best and fairest. But when we look back and forth through many decisions, a different picture emerges. No need to worry about minorities yet!—The majority is frustrated more often than not! Doesn't this destroy the original argument? For what was appealing about the satisfaction of the preference of the majority was precisely this: granted that all cannot be satisfied, the best is that most should be; the principle of letting the desire of each have equal weight will both secure this and in itself be fair, because none is given privileged consideration in the decisions as they come up.

But the table shows that the best and the fairest (by that criterion) may part company when we look through a multitude of decisions. However paradoxical it may be, it seems that the principle of the majority getting what they want allows the majority to get what they do *not* want in a majority of cases. The air of paradox is only partly produced by its looking as if some set of people, called 'the majority', were the same in both parts of the observation—as of course, it is not.

[1] May there not be some liberties which only stand so long as they are not brought into consideration?—Bringing them into consideration is *already* making them subject to decision.

That it is not only shows that what seemed paradoxical is not viciously so at all. It does not show that there is any error in the judgment: 'So in this way the majority may be frustrated!'

Note that this is not a point just about the method of decision by majority vote. I have been at pains to distinguish that from the principle of the decision's being what the majority wants. The method of majority decision may be regarded as *a* method for securing the latter. But the paradox to which I have pointed applies to the principle itself as well as to the method of majority decision.

It will now be plain why I have assumed the simplest possible kind of issues as the matter of decision, and assumed a perfect democracy in which all concerned vote quite directly. It is a familiar point that different voting procedures, multiple choices, what is motion and what is amendment, and other such matters, make such a difference to results that the will of the majority can seem to be uncapturable, or even non-existent. Also, we don't have pure, but rather delegated, democracies. One is inclined to say: 'If only everything were simple, the total population concerned voting on straight Yes–No alternatives without any question of rigging (by nature or design) from the order in which they were taken —if only things were like that, *then* we'd have the ideal satisfied (to the extent that people weren't stupid) of the greatest possible number of people getting what they want.' What the table shows is that this is an illusion. We ought therefore to give up any idea of achieving as good an approximation as possible to that ideally simple kind of result, which can itself be achieved only in very simple situations. Where the ideal is actually achieved the achievement of the further ideal is by no means guaranteed.

It may be, indeed, that that criterion of fairness appeals to us: the criterion by which the desire of each has equal weight in each decision. But when we look back and forth through many decisions, a different idea of fairness in satisfaction obtrudes itself. Thus a parent of a large family might say: 'Each child shall *take its turn* to decide such-and-such—*that* will be fair.'

Someone may say: 'The possibility which you have pointed to does not matter at all, even if it is often actualized. (And, by the way, you don't know whether it is often, never, or very seldom actualized.[1]) It is simply evident that majority decision is a sound way for a group of reasonable, equal, autonomous beings to determine what is best to do. That being so, it is of no import if, over many decisions, the method leads to this result. What matters is the determination of *each* question in the best way.'

To see both the force and the weakness of this comment, let us

[1] It would be of some interest, perhaps, to scrutinize the voting in the English Parliament of the early eighteenth century with just this question in mind.

imagine a method of majority judgment instead of majority decision. Our case will be rather artificial, but no matter. We will suppose some matter of fleeting observation of which constant record is needed, and where the observation always decides between a pair of alternatives. The matter being difficult, a team of observers is always employed, and what is put down in the record each time is the observation of the majority. This gives rise to just the same possibility as we have been discussing for decisions. When you look through a series of such results, you may find that the majority of the team have made the minority observation in a majority of the cases. But there would be nothing disturbing about this. If the method for determining what is to go in the record deserves criticism, that will not be because of *this* possibility; if the possibility is actualized, that is just a "giddy parergon", a merely incidental curiosity. It has no tendency to invalidate a "principle of majority observation". For all that matters is the determination of each result at the time. The same might hold if the principle of majority judgment were used in trials with a constant panel of judges.[1]

Such no doubt is the thought of our objector. He is saying: 'All that matters is the determination of each cause at the time in the best way'. But that can hardly be right, except on the assumption that there is an objectively best or good decision to make, the best way of determining which is to take a majority vote. By 'an objectively best or good decision' I mean a decision which is best or good independently of what people want. Well, if anyone can give such an account, more power to him! But as we have been considering the matter, all that is in question *is* what, given their circumstances, people actually want, which in the individual case a man gets by his choice or action if it is not stupid.[2] But in the case where people must do or suffer something collectively, the idea was that the majority should get what they want. So the frustration of the majority when we consider a whole series of majority decisions—the failure of the majority to have got what they want more often than not—cannot be regarded just as a mildly curious by-product.

Indeed, we might see in our problem a way of giving a real meaning to the rhetorically attractive but actually senseless expression 'the greatest happiness of the greatest number'. This is of the same form as 'the greatest number of words in the shortest time',[3] and when we have two things actually countable and measurable like words and times we see that it makes no sense. But we might say that what the famous phrase intends *can* be given a sense: at least it can if 'minimal frustration

[1] My colleague J. E. J. Altham has remarked to me that the smallest number for which this result is obtainable is five.

[2] This is not in fact true, because 'the race is not to the swift etc.' But the falsity of it cannot play a part in our considerations.

[3] The comparison was passed on to me by Gareth Evans, but I think he said he did not invent it.

of the majority' has one. How this could be planned for is indeed difficult to see. It might have to involve departing from the method of majority decision by the population.

That, of course, we do depart from anyway; we have delegated democracy, and use majority decision in committees and Parliaments. The aim of their decisions is not supposed to be satisfying the desires of the majority of their own membership. *Here,* then, the objection that I considered has a point; there is not necessarily anything to worry about, if the majority of a committee should turn out to have voted in the minority in the majority of issues it has decided.

Let us turn, then, to the question of satisfying the majority desire, on any given issue, of the whole of a community or population. As remarked, my table need not be a voting table: it may represent a set of questions being decided (no matter how) in accordance with, or contrary to, the will of people (or groups) A–K. Here then we see the naked point: the majority may be satisfied on every issue, while nevertheless the majority is frustrated over a majority of issues. There is thus the possibility of a certain technique of tyranny whose every measure has the support and is truly in accord with the desire of the majority, those whom any given measure hurts being in the minority; or again, one by one "merely sectional interests" are damaged. Since everyone not wretchedly isolated belongs to several "sections", it will be possible for the tyrant to damage the interests of anyone or any group (that does not support him, say) while truthfully claiming "democratic" support for his measures. Or again, the process of damage to sectional interests—that is, to a majority of the population—may occur in a democracy in a haphazard fashion and without design, always in accordance with the will of the majority.

University of Cambridge

EGOISTIC HEDONISM

By Lawrence Crocker

DESIRES differ from each other along several dimensions. They differ in intensity and probably in a hundred distinct ways with respect to their phenomenological felt character. Some desires are in part linguistic in character—as when we desire something under a certain description. Other desires seem to have no linguistic component. For the purposes of this paper I will abstract from most of these ways

that desires differ and distinguish them primarily along two dimensions —their objects and whether those objects are desired as means only or at least partly as ends.

Let an 'end-desire' be a desire for something as an end, an 'S-desire' an end-desire for a state of the desiring individual's own consciousness and an 'O-desire' an end-desire which is not an S-desire. I want to consider two claims

(P) All end-desires are S-desires

(E) It is irrational to pay a cost for the fulfilment of an O-desire.

I

(P) is a liberalized statement of psychological egoistic hedonism—liberalized in that it does not require that the states of consciousness be pleasurable. It has frequently been observed that psychological egoistic hedonism, even liberalized in this fashion, must be false since we obviously do want good things for other people. But it has always been necessary to defend this observation against the contention that we deceive ourselves into believing that these altruistic desires are end-desires. I think that I want you to enjoy a good dinner, but really I want you to enjoy a good dinner which I have cooked, so as to put you in my debt.

It does seem that we sometimes deceive ourselves about our "unselfish" desires in this way. But the contention that there are no altruistic end-desires[1] is implausible, even though some philosophers and not a few ordinary sensible people believe it. As has often been pointed out, it flies in the face of certain facts, such as self-sacrifice and the preparation of wills. Moreover desires for the welfare of other members of the species, in particular those with which the individual is in close contact, and especially the young, have considerable survival value. It is a trait widely distributed among the higher animals, and it seems not unlikely that it was shared by our remote ancestors, with whom we are genetically identical.

In short, the contention that we do not really have altruistic desires is not particularly plausible and, in any event, requires more evidence than has been given by those who argue for some version of (P). (Surely that is where the burden of proof lies, since so many people think that they

[1] Notice that this claim is stronger than the contention that all our altruistic desires are themselves means to another end. A desire can be a means to an end while itself being an end-desire. (Imagine that someone offers you a million dollars if you, in the next ten minutes, have a desire for fresh strawberries. You will try to cultivate the desire, probably by thinking of the gustatory attributes of fresh strawberries. Your *wanting* the strawberries is a means to a million dollars, but you do not want the *strawberries* as a means to a million dollars. The strawberries aren't even part of the means to the million. Having the strawberries makes you not one whit closer to the million. Nor will failing to get the strawberries keep you poor.)

have altruistic desires.) But my intention in this paper is to make a case against (P), and against (E), *without* relying on the phenomenon of altruism.

II

It may seem at first as if (P) can be cleanly defeated by reference to something as mundane as a desire for a dish of ice cream. Consider an occasion on which I know that my nutritional needs for the day have been more than adequately met, so that the ice cream is not a means of that sort. Then what I seem to have is a single, complex end-desire. In part it is an S-desire. Primarily this is a desire for the taste of the ice cream. Also involved are desires for sensations having to do with the temperature, texture and appearance of the ice cream. But what is striking is that the object of my desire is not exhausted by these sensations. In fact it takes a mental exercise just to isolate the sensations. I don't desire a sensation of taste; what I desire is to eat ice cream. The object of the desire is, in fact, an event in the physical world—involving a certain substance and certain actions as well as certain sensations. If, then, the desire in question is an end-desire, it certainly has an O-component and (P) is false.

But there is an alternative account: What seems on first glance to be a single end-desire is, in fact, in part a desire for some of its objects as means. The ice cream, its transportation from dish to mouth, *etc.*, are all desired as means to the sensations. If this is right, the only genuine end-desire is an S-desire.

One might try to rule out this second interpretation on the basis of the following consideration: Since the means-end relation is a matter of calculation, an apparently unitary desire cannot be divided into means-components and end-components. To be a means something must be *thought of* as a means. If this objection were telling, it would make the original account of a desire for ice cream nearly mandatory. (It would also form a bulwark in the defence of the existence of altruistic end-desires.) But the objection seems to be mistaken. The calculational nature of means and ends is apparently compatible with subconscious means-desires. We carry out all sorts of complex calculational tasks subconsciously. Moreover, it is relatively rare that we consciously think something of the form 'I want x as a means to y'. Only occasionally do I think of putting on my shoes or coming to my office as a means to doing something else. Yet these are means—means whose ends I can specify with only a little reflection. In general it seems that a large percentage of the behaviour that must be explained in means-end form is not consciously thought of in that way when it is performed.

Eating ice cream is a means to taste sensations. That fact does not entail that I desire to eat ice cream solely as a means to those sensations.

But neither does the fact that I do not consciously distinguish means and end components of my desire show that I do not desire to eat ice cream solely as a means to the sensations.

At first it might seem as if the crucial test would be to ask whether I would desire to eat ice cream if it were not a means to the sensations of tasting ice cream. An affirmative answer is sufficient to show that the desire to eat ice cream has an O-component. But it is not necessary. One may have a desire for something as an end when it is part of a certain context which one would not have for it in isolation. The better question is this: Would I desire the sensations of tasting ice cream any less if they were the result of some sort of simulation rather than the usual process of eating ice cream? In my own case I don't think that I would. Perhaps there are people who would answer the question in the other way—who, nutrition aside, would desire real ice cream more than a sensationally perfect counterfeit. But it is possible that such persons are self-deceived or unable to take the terms of the case seriously. They may be influenced by the belief that there could be no perfect counterfeit of eating ice cream or the belief that any such counterfeit would probably be unhealthy. Why otherwise should they prefer the real thing? This last challenge to the fans of genuine ice cream is not, of course, decisive, even if unanswerable. We are not yet considering what desires we *should* have, in any sense of 'should', but simply what desires we do have.

It may seem as if a better counterexample to (P) can be found among the desires for major consumer goods, paradigmatically the new car. I suspect that such desires frequently are not exhausted by component desires for all the satisfactions that fulfilment of the desire can be expected to bring. But the analysis of the desire for a new car is made difficult by the following complications: First, the desire may include component desires for ends which are fantasized (perhaps semi-consciously) and not reasonably expectable. Second, ownership of the car rather than its use may be a means to some of the satisfactions. Third, such desires sometimes have a peculiar self-intensifying character (which in its extreme form produces obsession). The first two of these facts weakens the case as a counterexample to (P). The third may strengthen or weaken the case, depending on the way it is analysed. Because of such complexities, desires like those for major consumer goods fail to give rise to clear and decisive arguments against (P).

III

A stronger case against (P) can be made by considering the desire to be loved. Suppose that the angel of possible worlds presents me with a choice of which of two very similar possible worlds to live in. In each world I have exactly the same states of consciousness. The difference

between the two worlds is that in one of the worlds I am loved while in the other I merely have all the conscious states which accompany being loved, including the well confirmed belief that I am loved. The angel asks me in which world I would prefer to live. I would reply that I would prefer to live in the world in which I was loved, even though it would be no more pleasant than the other world.

My response is a *prima facie* counterinstance to (P). It is only *prima facie* since there is, again, the possibility that my judgment with respect to this somewhat non-ordinary case does not reflect the true state of my desires. It may be that my judgment is put off track by unsuppressible knowledge that it is impossible to have *all* the states of consciousness which accompany being loved unless one is loved, or that if one is systematically deceived there is *always* a possibility of discovery. I doubt that I am led astray in one of these ways since my intuition doesn't waver when I emphasize to myself the conditions of the case. Perhaps I am misleading myself, but a demonstration to that effect could hardly consist of anything short of strong independent evidence for the truth of psychological egoistic hedonism. The burden seems to rest squarely on the supporters of (P) to produce such evidence.

IV

I am satisfied, on the basis of this case, that (P) is false. I have been troubled, however, about (E) which is either a form of or fairly closely related to ethical egoistic hedonism.

Suppose that the angel asks me what cost I would pay, on reflection, to live in the world in which I really am loved as against the world in which I merely have all the appearances. Cost is specified as follows: in each world my life is made up of pleasant days, in fact, days in all respects as good as a specified good day last week. The cost is the elimination of days from the end of my life in the world I have chosen. It would not be very easy to answer the angel's question in these terms, but my initial intuitions are that I would put in a bid of more than one day but not of more than two years. Here, then, on the basis of at least some reflection I would be willing to pay a cost for the fulfilment of an O-component.

But is it really *rational* to pay a cost to live in one of these worlds rather than the other? I have intuitions in both directions. I don't see why I should disown this desire any more than any other desire, but I have the nagging feeling that perhaps I am being a chump if I pay a cost for something which would have no effect on my consciousness. The remainder of this essay will explore whether there is any substance to this negative intuition, since the burden of proof seems to lie on the negative side.

V

The consideration which accompanies the initial negative intuition is the observation that the difference between the two worlds doesn't make any difference *to me*. But in what sense doesn't it make a difference to me? There is the difference that in the one world I get what I want, I am loved, while in the other world I do not get what I want. Of course the subjective side of the fulfilment for me is the same in the two worlds, but to characterize 'difference to me' in terms of subjective states is to beg the question.

It might be contended that the difference between the two worlds is of no *benefit* to me. But this way of putting the matter is of no help. The only reason not to count the difference between true and simulated love as a benefit is an arbitrary restriction of benefits to states of consciousness. With that restriction the point at dispute is simply restated without gain as 'Is it rational to pay a cost for something which is of no benefit to me?'

VI

Can anything akin to a traditional principle of reason be used to disqualify paying a cost to live in the preferred world? There seem to be only two principles which are remote possibilities. One should not pay any cost for the fulfilment of a desire if it

1. Conflicts with the fulfilment of a desire more favoured on reflection

2. Is composed of parts, the fulfilment of none of which one would pay a cost for on reflection.

(1) is clearly a sound principle, but it does not apply in the present case since the only conflict with other desires is the cost I am considering paying. (2) is unsound as it stands since one can have a desire for a whole on reflection, without desiring any of the parts in isolation on reflection. Even if some amended version of this principle is sound, it presumably would not apply to the present case since the O-component of the desire to be loved does seem to have desirable parts.

VII

It is a strong empirical conjecture, but not an entirely implausible one, that the O-component of an end-desire comes to accompany a core S-desire as the result of reinforcement. If someone really has an O-desire for ice cream that may be solely because eating ice cream has been accompanied by pleasant sensations. Similarly I may come to desire to be loved solely because in the past I have been happy when I believed that I was loved. If positive reinforcement is the whole explanation for the existence of O-desires, one might make the following argument in

favour of (E): Reinforcement mimics means-end calculation in that it produces a desire for something which may reasonably be thought to have desirable effects. But the mimicking is imperfect both because reinforcement doesn't require any real connection between what is reinforced and the reinforcers and because the produced desire is for the thing not just as a means but as an end. The irrationality of O-desires consists in their origin in this distortion of means-end calculation.

The chief problem with this argument, its strong premiss aside, is that the origin of a desire may be irrational—or non-rational—without its being irrational to pay a cost for the fulfilment of that desire. Most desires, including fundamental ones, arise in ways that are even less similar to a rational mechanism than is reinforcement. An end-desire, unlike a means-desire, is just not the sort of thing for which a rational origin or justification should be expected.

VIII

Still if the strong reinforcement hypothesis were true, there would be two considerations which could, arguably, be used to deflate the cost one should be willing to pay for an O-desire. The starting point of the first is the fact that the origin of such a desire is, in part, external. Consider a desire whose origin is entirely external—for example, a desire I come to have as the result of a hypnotic suggestion. In this case I could reason that, though I have the desire, there is a sense in which it is not really *my* desire. It seems reasonable that I should pay less for the fulfilment of a desire which is not in all respects mine than for one which is. Possibly a case can be made for the claim that all O-desires are externally imposed in at least an extended sense—if not by reinforcement alone, then with the help of persuasion, group pressure and second order desires (perhaps initiated by a desire to be the sort of person who would have a given O-desire). Whether these are sufficiently "external" mechanisms and whether it is plausible that together they account for all O-desires are questions that can be passed over here. Even if the answers to both were affirmative, it would not be sufficient to reduce to zero the cost I should pay for the fulfilment of an O-desire. If I have a desire, whatever its origins, it is in some respects my desire. If it survives (1) above, I don't yet see why I shouldn't pay some cost for its fulfilment, even if I have it as the result of hypnotic suggestion. There is the possibility, of course, that I may cease to have a desire altogether when I discover its origins. But that certainly wouldn't happen in every case. Suspecting that desires are the result of reinforcement, in particular, does not usually cause them to disappear.

The acquisition of O-desires through reinforcement or other "external" mechanisms might also suggest that such desires were superficial, though desires of internal origin might also be superficial in

the sense of not being very deep in the personality structure. The idea here is that a desire is superficial if and only if I could fail to have it without being very different from what I am. Superficiality in this sense is presumably compatible with a fairly high level of intensity. If this notion of the depth of a desire makes sense, and if the O-desire to be loved turns out to be superficial (which seems unlikely), that would, again, be a reason to reduce my bid to the angel of possible worlds, but I fail to see how it could be a reason to reduce that bid to zero.

IX

Perhaps the most plausible way to argue for (E) is to contend that there is no essential difference between the perfect simulation of an ice cream experience and the perfect simulation of love. Such an argument, if successful, would force me to give up my counterexample to (E), since I am more certain of my unwillingness on reflection to pay a cost for true ice cream (as against the perfect simulation) than I am of my willingness to pay a cost for true love.

Unfortunately for this line of argument, there does seem to be a significant difference between the two cases. It comes out when one asks how much less desire one would have for known simulation—say knowing that one has electrodes in one's head in the one case, or knowing that one is faced with a consummate actress or actor in the other. After overcoming an initial aversion to electrodes, I think that I would be indifferent whether I was having a real or a simulated ice cream experience. Certainly I would prefer ice cream simulation known rather than unknown since there is something slightly pathetic about the individual who thinks he is eating ice cream, but is really having his brain stimulated. The individual who has his brain stimulated knowingly may not win any friends in the natural foods crowd, but he is not pathetic. On the other hand, knowing that love is simulated eliminates most or all of its charm. So, in contrast to ice cream, true love is preferable to simulated love, and simulated love unknown is preferable to simulated love known.

The reason for this is, I think, that being loved, like being admired or respected, is an ingredient in the individual's conception of his or her own worth. In this respect it is a kind of excellence like being accomplished, virtuous or knowing the truth. Desires for excellences, in general, seem to survive reflection quite well. In fact I am tempted to say that the only sorts of desires that normally do prove on reflective consideration to be continuing end-desires are desires for states of consciousness (one's own and others'), for excellences (one's own and others') and certain second order desires. In any event, it appears that there is here a distinction which makes a difference. Not to have a

genuine ice cream experience would not affect my estimate of my worth; not to be genuinely loved would.[1]

Of course in both worlds I will have self esteem. But in only one of them will it be justified, and that is what I want. One can still ask why one *should* want to have an excellence rather than just the appearance of an excellence. But why shouldn't one? It seems perfectly natural to care about one's worth, rather than just its appearance. In fact the latter would seem at least to border on hypocrisy. In the end the burden of proof is again on the side of the supporters of (E) to show that one shouldn't care about really having an excellence and, thus, that this difference between the two cases shouldn't make a difference.

X

Since (P) and (E) both seem insupportable, where do the intuitions in their favour come from? In part it may be a matter of such conceptual mistakes as those of section V. But I doubt that views are ever as appealing as these views are solely as the result of conceptual confusions. The single most important cause of the popularity of (P) and (E) is, I think, the overwhelmingly obvious fact that for each of us our own states of consciousness are right at the centre of our system of motivations and desires. We are so constantly and so intensely concerned with our own states of consciousness, that it is easy to suspect that perhaps we are, or should be, concerned with them alone. This suspicion grows with the discovery that not everything which seems to be an O-desire really is. There may be more cases where apparent O-desires vanish under reflection than cases where we are willing to pay a cost for their fulfilment. In addition, some O-desires may be externally induced or superficial. Taken uncritically this is all evidence for either (P) or (E) or both. Of course to get (P) or (E) from this evidence one must overgeneralize. But there is good independent reason to believe that overgeneralization is a failing to which our intuitions are chronically liable.

University of Washington

[1] Desires actively to accomplish something provide perhaps even clearer cases of the ineliminability of O-components which are desires for excellences. The angel of possible worlds can give me an experience subjectively identical to that of climbing Mt. McKinley without its being an experience of climbing Mt. McKinley. But it is the latter which I want, because I want to succeed at actually climbing the mountain.

Compare this with a desire for the existence of G. E. Moore's beautiful planet which no one will ever see. This desire is a desire for an excellence—not of a person, but of the universe. I am inclined to think that this makes a crucial difference. Even if one had the desire that Moore's planet should exist (unsuspected), it would be irrational to pay a cost for the fulfilment of this desire.

HARE ON ABORTION

By RICHARD WERNER

IN R. M. Hare's recent article 'Abortion and the Golden Rule',[1] he argues that too much effort has been wasted on determining the ontological status of the unborn. Indeed Hare believes that the ontological issue cannot be separated from the moral one and, accordingly, that we should just get down to the business of dealing with the moral question. His reasoning appears to go something like this

1. 'If a normative or evaluative principle is framed in terms of a predicate which has fuzzy edges (as nearly all predicates in practice have) then we are not going to be able to use the principle to decide cases on the borderline without doing some normation or evaluation' (p 204).
2. '"Person", even if descriptive, is not a fully determined concept; it is loose at the edges, as the abortion controversy only too clearly shows' (p 205).
3. The unborn lie on the fuzzy edges of the concept of being a person (suppressed premiss).
4. Hence any normative or evaluative principle which is framed in terms of the concept of a person cannot be used to decide cases concerning the unborn without doing some normation or evaluation (p 205).
5. 'It is not necessary, to insist that the word "person" is a moral word', to show that deciding the ontological question is a moral decision (p 205).
6. 'If we decide that, "within the meaning of" the principle about murder, a fetus becomes a person as soon as it is conceived, we are deciding a moral question' (p 205).
7. Therefore, given both that deciding the ontological question is a normative decision and that within the meaning of the principle about murder it will decide the moral issue of abortion, determining the ontological status of the unborn must be a moral decision (p 205).

As a result of this argument, Hare asserts that 'the reasons he [the anti-abortionist] will have to give in saying that it [the unborn] is a person and that, therefore, killing it is wrong (or that it is not a person and, therefore, killing it is not wrong) will be the very same moral reasons as I shall be giving to my more direct question' (p 205), namely, 'How ought a creature about whose properties, circumstances and probable future we are quite adequately informed, to be treated?'

[1] *Philosophy and Public Affairs*, vol. 4, no. 3 (Spring 1975): 201–222. All page references contained within the body of this paper are to Hare's article.

(p 204). 'Whichever way one takes it, one cannot avoid giving a reasoned answer to the moral question; so why not take it the simplest way?' (p 206)—i.e. Hare's way of resting the potentiality principle on his own universal prescriptivism.

There are several problems I would like to raise with Hare's argument. First, premiss 3 of the argument is a central premiss, yet nowhere within the article does Hare mention or defend it. Secondly, this premiss is controversial, and does not in fact follow from premiss 2, as I shall attempt to show with the help of the following distinction.

By 'human being' I mean a member of the biological species *homo sapiens*. By 'person' I mean a fully fledged member of a human community, someone having a developed concept of *self*, memories, a language and/or moral obligations as well as moral rights (if one wants to keep the notion of a person morally neutral then ignore this last criterion). This is the sense of 'person' which philosophers are most interested in capturing when they deal with such issues as personal identity and the problem of other minds. I make this distinction because I believe it to be useful in avoiding confusions and ambiguities when discussing such issues as abortion. Further, it is a distinction I believe to be warranted by critical common sense. The justification is a pragmatic one and does not, as far as I can see, beg any questions its refusal would not equally beg.

Hare might argue that the decision to accept this distinction between persons and humans is a normative decision having moral implications and, as such, is a moral decision needing moral justification. If that is the case, then it would be interesting to know why decisions to accept distinctions such as those between the moral and non-moral, the cognitive and the noncognitive, the normative and the descriptive and the prescriptive and the non-prescriptive—distinctions all central to Hare's supposedly morally neutral meta-ethical position—are not equally moral decisions standing in need of moral justification. Clearly, these distinctions have far-reaching implications for Hare's universal prescriptivism and his resulting moral prescriptions. Furthermore, these concepts are as fuzzy edged as any predicate is likely to be 'in practice'. Hare cannot bring this charge against our distinction without being hoist by his own petard.

Now let me explain my remarks concerning premiss 3. One may be willing to grant that 'human being', in its biological sense, or 'person', in its philosophical or ordinary sense, are cluster concepts and, as such, have fuzzy boundaries. However, in the same breath, one may deny that the unborn lie on the unclear boundaries of either of these two concepts. If this is correct, then Hare's argument fails to shown that there is anything importantly evaluative or moral in using these concepts to determine the abortion issue. For instance, one may grant that certain

cases of retarded humans, or strangely different but still similar aliens from another planet, fall on the unclear boundaries of the concept of being a person. Likewise one may admit that if monkeys throughout the world began to give birth to creatures which looked and behaved more like humans than monkeys, or if women throughout the world began giving birth to creatures which looked and behaved more like monkeys than humans, these creatures would fall on the unclear boundaries of the concept *human*. But this is perfectly compatible with denying that the unborn fall on either of these fuzzy edges. The unborn are clearly not persons in either the ordinary or the philosophical sense, they have no concept of *self*, no memories, no language, no moral obligations; but clearly they are human beings, members of the biological species *homo sapiens*. The unborn are the live, growing, uterine offspring of humans and are not members of any other plant or animal species. Nor can they be passed off as extra organs, biological parts or limbs belonging to the mother's body. At the very least, then, there is very good reason to doubt, and Hare has certainly failed to establish, the truth of premiss 3. Without it, his argument cannot even begin to get off the ground.

Hare's second main problem is in attempting to move from premiss 4 to his conclusion 7. For, if our preceding analysis is correct, then 4 does not follow and Hare has not established that there is anything importantly normative about the use of the concept of a person in deciding issues concerning the unborn. As such, we need not beg any moral questions in deciding this issue. It could be done on non-moral grounds. The conclusion, 7, does not follow. If there is nothing importantly normative about the use of the concepts of a person or human being within the abortion controversy, coming to a decision on this issue of the ontological status of the unborn is no different in kind from determining a driver's speed prior to an automobile accident so as to determine the driver's legal or moral culpability. Neither need be a moral or question-begging matter.

Further, it will not do to argue that great disagreement over the ontological issue shows it to be a normative one. There has been great disagreement over many factual issues—the ontological status of racial and ethnic minorities, the shape of the earth, the relationship between I.Q. and heredity, the viability of Copernican astronomy and quantum mechanics. As Hare indicates, the issue here is a philosophical one. But it is an issue which requires properly interpreting and analysing the facts at hand, rather than inventing new beginnings or discovering new facts. Indeed, much of the disagreement over the ontological status of the unborn arises, I believe, because of our failure to distinguish clearly between human beings and persons rather than through any inherent inscrutability in the application of these notions to the abortion issue.

The concepts of a person and a human being may both have very fuzzy boundaries; there may be great disagreement over the ontological status of the unborn; and still the ontological issue of the unborn may be a non-moral, philosophical one.[1]

Hare might attempt to counter this by pointing out that our distinction between persons and humans only pushes the normative, moral question one step further back. How we decide the question 'Should persons or humans have a full right to life?' is a normative issue. Since our decision will settle, or at least begin to settle, the moral issue of abortion, it will also be a moral decision. Accordingly, all that Hare says following 7 applies, and his argument is saved.

I doubt if anyone would want to deny that the question 'Should human beings or persons have a full right to life?' is a question demanding a moral answer. The 'should' here is clearly a moral one, as appears particularly when we realize that our answer will help settle such issues as the moral legitimacy of abortion, infanticide, human experimentation, the rights of future generations and so on. But in settling this issue, Hare's universal prescriptivism is of no help whatsoever. 'We should do to others what we are glad was done to us', does not and cannot help us in determining whether full moral rights are to be granted to human beings, persons or both. However we interpret the 'we', 'others' and 'us' in Hare's principle, whether we take these terms as denoting humans or persons, we will beg the moral question as to whether persons or humans are to have full moral rights. As such, deciding whether human beings or persons are to have full moral rights will have to be done on the same grounds as those used prior to Hare's article, i.e. by attempting to determine which criterion best captures our critical common sense beliefs concerning such issues as infanticide, the killing of the deformed and retarded, experimentation on children and the unborn and the rights of future generations.[2]

Since, as we have seen, there need be nothing particularly normative or moral about deciding the ontological status of the unborn in the abortion issue, Hare's justification for bringing in the potentiality principle, based on universal prescriptivism, is unwarranted. Any appeal to such a potentiality principle will have to be justified on non-moral grounds. Accordingly, even if we grant that deciding whether persons or humans have full moral rights is a moral question, nothing else of what Hare has stated follows.

[1] See M. A. Warren, 'Moral and Legal Status of Abortion', *The Monist*, vol. 57, no. 1 (Jan. 1973) and M. Tooley, 'Abortion and Infanticide', *Philosophy and Public Affairs*, vol. 2, no. 1 (Fall 1972) and my 'Abortion: The Moral Status of the Unborn', *Social Theory and Practice*, vol. 3, no. 2 (Fall 1974) for a more detailed discussion of how this distinction helps to clarify matters in the abortion issue.

[2] See Warren and Tooley, *opp. citt.*, who argue in favour of taking personality as the morally relevant criterion. See Tooley correspondence, *Philosophy and Public Affairs*, vol. 2, no. 4 (Summer 1973) and my article, *op. cit.*, for the opposing point of view.

So, in conclusion, I have adduced two major objections to Hare's position. The first is designed to show that there is no need to assume that deciding the ontological status of the unborn begs any normative questions. The second is intended to show that, if this be so, there is nothing particularly moral about determining the ontological status of the unborn and that Hare's move to bring in the potentiality principle based on universal prescriptivism is therefore unwarranted. These criticisms, if correct, render Hare's argument unsound.

Kirkland College

PRIOR ON PROPOSITIONAL IDENTITY

By PHILIP HUGLY *and* CHARLES SAYWARD

IN a natural language we say such things as that John and Bill agreed on some points. This puts us in mind of the possibility of constructing a logic incorporating what we may call *propositional quantification.*

Reflection on the relevant sentences of a natural language may then well lead us to think of a sentence like 'John believes that Bill is happy' as containing 'believes' as a relational predicate, flanked to the left by 'John', taken to be a name of John, and to the right by 'that Bill is happy', taken to be a name of a certain special object: the proposition that Bill is happy. Granting this analysis, the sentence 'John believes something' goes over into

$\exists x$ John believes x

where 'x' is thought of as a name-variable in referential position. In a system of logic erected on the basis of this kind of analysis, the laws of propositional identity will just be those of ordinary object identity:

(1) $\forall x\, x = x$
(2) $\forall x \forall y\, (x = y\, .\, \supset\, .\, Fx \supset Fy)$

From (1) and (2)

(3) $\forall x \forall y\, (x = y\, .\, \supset\, .\, f(x) = f(y))$

is derivable, where '$f(\,)$' is a functor forming names from names. But the converse of (3), namely

(4) $\forall x \forall y\, (f(x) = f(y)\, .\, \supset\, .\, x = y)$

does not hold, e.g., where 'f' abbreviates such a functor as 'the proposition John asserted just before he asserted . . .'.

In *Objects of Thought* (Geach and Kenny, eds., Oxford University Press, 1971) Prior attempts to develop a logic for propositional quantification. But he rejects the kind of analysis sketched above. Rather he analyses the sentence 'John believes that Bill is happy' into the operator 'believes that' flanked to the left by the name 'John' and to the right by the *sentence* 'Bill is happy', taking it that sentences are not names, so that their positions are non-referential. On this analysis the sentence 'John believes something' goes over into

$\exists p$ John believes that p

where 'p' is an essentially bindable sentence-letter in non-referential position. This conception of propositional quantification eschews the notion that propositions are a kind of object, or subject to any kind of

naming, and gives to the quantifier '∃' the reading 'some' rather than the reading 'there exists'.

Now on page 101 Prior sets out a logical system for propositional quantification and identity having the following features:

> . . . (a) ordinary propositional calculus, enriched with variables for expressions which form sentences from sentences, with quantifiers binding variables standing for sentences, and with the identity function with sentences as arguments; (b) the ordinary theory of quantification applied to our special quantifiers; and (c) ordinary laws of identity applied to our special function.

The resulting system (call it '*S*') is rich enough to yield interesting results. For example, Prior uses it to give a particularly ingenious treatment of Moore's Paradox (*Objects of Thought*, pp. 79–83). But it is worth noting that Prior thinks that

(5) $\forall p\ p = p$
(6) $\forall p \forall q\ (p = q \supset \delta p = \delta q)$

are the only axioms needed for propositional identity. What is striking here is that (5) and (6) are entirely parallel to (1) and (3) though in the latter '*x*' and '*y*' are name-variables in referential position rather than sentence-letters in non-referential position and '*f*' stands for operators forming names from names whereas 'δ' stands for operators forming sentences from sentences. Since sentences are not names, an analogue of (2) which kept '*F*' to a predicative role would be ill formed in *S*; only an analogue of (2) which gave '*F*' the role of a sentential operator would be well formed in *S*, and in that case the analogue would be a mere stylistic variant of (6).

In fact, Prior has not seen a significant disanalogy between the way '=' operates in the case of individual variables and the way it operates in the case of sentence letters. In the former case, (4) is not a logical truth. In fact, for any distinct objects *a* and *b* one can always come up with some function *f* such that $f(a) = f(b)$. However, in the case of sentence letters (where 'δ' is a variable standing in for operators which form sentences from sentences)

(7) $\forall p \forall q\ (\delta p = \delta q\ .\ \supset\ .\ p = q)$

is arguably a logical truth. And, in any case, various instances of (7) without doubt *are* logical truths: if the proposition that $\sim p$ is the same as the proposition that $\sim q$ (which, following Prior, is a fluffed up way of saying $\sim p = \sim q$), then the proposition that *p* must be identical with the proposition that *q* (i.e. $p = q$); if $(p \vee r) = (q \vee r)$, then $p = q$; and so on.

As long as these instances of (7) are admitted as logical truths, it follows that feature (c) does not suffice for propositional identity: for

neither (7) nor any of the instances of (7) are theses of S. So *some* strengthening of S is called for.

If (7) itself is a logical truth, then, since (7), (5) and (6) are independent, this strengthening seems best achieved by adding (7) as an axiom of S. On the other hand, if (7) is not a logical truth, it is not at all obvious how S is to be suitably strengthened.

In effect, it seems, where 'δ' stands for any operator which forms *sentences* from *sentences* (e.g., 'not', 'John believes that', etc.) $\ulcorner \delta p \urcorner$ expresses the same proposition as does $\ulcorner \delta q \urcorner$ only if p and q express the same proposition. This seems to be a fundamental difference between those operators which form sentences from sentences and those which form names from names and this difference must find expression in a system of propositional quantification which, like Prior's, holds propositional variables to sentence position and introduces an identity sign as a sentence connective.

The following considerations give rise to potential counterexamples to (7). Suppose we put for 'δ' in (7) an operator '*' such that for any sentence p of S $\ulcorner *p \urcorner$ is a tautology. Then, if tautologous propositions are identical

$$(8)\quad \exists p \exists q\, ((*p = *q)\ \&\ (p \neq q))$$

will be true, and (7) false. The tenability of this depends on whether it is correct that any two tautologous propositions are identical. It may well be that general theoretical considerations should dictate how one decides such an issue. If so, the considerations brought forward two paragraphs above seem to us to provide powerful reasons for rejecting the view that all tautologous propositions are identical and for accepting (7) as a logical truth.

University of Nebraska—Lincoln

DAVIDSON'S CRITERION OF EVENT IDENTITY

By J. E. TILES

IF 'event' is a sortal term at all, it is, I suspect, what David Wiggins (p. 29) called a 'dummy sortal'. That is to say, there is no single criterion of identity for everything that may be called 'an event'. Donald Davidson thinks otherwise, and in Davidson (2), p. 231, has offered the following criterion for event identity (where x and y are events)

> $x = y$ if and only if $((z)$ (z caused $x \rightarrow z$ caused y) & (z) (x caused $z \rightarrow y$ caused z)).

Davidson evidently expects initial reaction to this to be that of dissatisfaction, for he hurries to point out that this is not *formally* circular, since no identities appear on the right hand side of the biconditional. A proposed criterion, however, may be unusable for reasons other than formal circularity.

If this criterion is meant to persuade us that we know what we are talking about when we speak of (Davidsonian) events, then we should remain unpersuaded because: (1) It is unusable in the absence of a formal theory of *cause*. (2) It is unusable because we have not been placed in a position to distinguish events from non-events. (3) If a plausible thesis about the cause relation (to which Davidson himself once subscribed) is true, it is unusable because determining whether the *cause* relation holds between events A and B requires questions of event identity already to have been settled.

(1) Presenting identity criteria as definitions of the identity relation has a precedent in Frege's consideration of contextual definitions in *Grundlagen* §§ 55–69. Although Frege ultimately rejected the use of such definitions, there are two respects in which it might be felt that such definitions are superior to that which Davidson offers. For example, Frege considers defining the concept *the direction of a line*

> The direction of line a = the direction of line b if and only if line a is parallel to line b.

Not only does the identity sign fail to appear on the right hand side of the biconditional, but *neither* a singular term or variable mentioning directions *nor* a quantifier whose range includes directions appears on the right hand side. The right hand side of Davidson's definition mentions the events between which the identity relation is to be defined (i.e. x and y) and quantifies ('for all z') over a range which includes events.

If one specifies a real number by quantifying over real numbers (a domain including that number) one is said to have given an impredicative specification of that number. This procedure—once called into question by Russell and Poincaré—is now widely accepted by mathematicians and logicians as innocent of "vicious circularity". Davidson, however, is not attempting to specify a particular event by quantifying over a range including that event, but is offering a general criterion of identity for the objects in the domain of events. If one requires the identity criterion for the objects of a domain to be fixed before quantifying over that domain, then Davidson does seem open to the charge of circularity.

In his own defence Davidson would probably point to the axiom of extensionality in set theory as the model of his proposed criterion

$$x = y \text{ if and only if } (z)\,(z \in x \longleftrightarrow z \in y).$$

Here both features which were sources of complaint in the previous paragraph are present, nevertheless the axiom is widely accepted as a definition of set identity. But in, for example, ZF set theory, this axiom is accompanied by a number of other axioms which supply a substantial part of the relevant properties of the $\in$ relation. Davidson would here be relying on the possibility of a formal analysis of the *cause* relation to underwrite his proposed criterion. (One thing which a formal treatment of the *cause* relation would need to explain is why it is sufficient for Davidson to express his criterion with conditionals and not biconditionals on the right of the major connective.)

(2) The axiom of extensionality allows only one memberless object in the universe of ZF, namely the empty set. To create a more familiar kind of set theory out of ZF, one has to restrict the x and y of the axiom to sets and allow the quantifier, 'for all z', to range more widely than sets. In this case the axiom relies on the criteria of identity of those objects in the range of 'for all z' which are not sets. It would appear that Davidson's criterion may be taken more in this spirit, for he does not say that the quantifier, 'for all z', is restricted to events.

This means that we cannot confidently turn to the range of significance of the *cause* relation to determine more precisely what Davidson would be prepared to label as an event. States are said to be causes (the car skidded because the tyres were worn) and to be effects (the tyres were worn by 100,000 miles of travel). Are states to be considered as events, their identity criterion given by Davidson's formula, or do they require a separate account? People are said to be causes. Are we to go so far as to count people as events? Clearly we have to know what else is in the range of the quantifier, 'for all z', and how to distinguish these other objects from events, if we are to make any use of Davidson's criterion.

(3) Davidson denies he is claiming that 'the only way of establishing, or supporting, a claim that two events are identical is by giving causal evidence' ((2), p. 231). Nevertheless he does hold it is a *sufficient* criterion (ibid.), which I take to be the claim that given two event descriptions it is always possible to determine on the basis of causal facts alone (given enough of them) whether they are of identical or distinct events.

But if we take into account a point made in a rough characterization of the 'analysis of singular causal statements hinted at' in another article by Davidson, we find ourselves caught in a circle which, however non-formal, seems thoroughly vicious:

> 'A caused B' is true if and only if there are descriptions of A and B such that the sentence obtained by putting these descriptions for 'A' and 'B' in 'A caused B' follows from a true causal law. (Davidson (1), p. 92n.)

Our judgment about whether A caused B will, therefore, rest on judgments of the identity of events. For to redescribe A and B, say as C and D (where we have a causal law which yields 'C caused D') is to judge A is identical to C and B identical to D.

If these latter judgments depend on whether A and C, B and D, have the same causes and effects, consider our position: In the absence of a true causal law yielding 'C caused B', we are in no position to say whether B and D satisfy one condition of identity, namely having the same cause C. Suppose we are fortunate enough to have such a law. Then an affirmative judgment that A is identical to C requires that A and C have the same effects. One effect of C is now known to be B, so we must determine if B is an effect of A. But this is precisely what we are trying to settle by determining whether A is identical to C.

Formal circularity in a definition or criterion is undesirable precisely because it renders the definition or criterion unusable. Even if Davidson is right in claiming his criterion is not formally circular, it is, as it stands, of no more worth than if it had been.

University of Reading

REFERENCES

Davidson, Donald (1)—'Actions, Reasons, and Causes' (1963), reprinted in Alan R. White (ed.), *The Philosophy of Action*, Oxford, Oxford University Press, 1968.
Davidson, Donald (2)—'The Individuation of Events' in N. Rescher *et al.* (eds.), *Essays in Honor of Carl G. Hempel*, Dordrecht, D. Reidel, 1969.
Frege, Gottlob—*The Foundations of Arithmetic*, Oxford, Blackwell, 1959.
Wiggins, David—*Identity and Spatio-Temporal Continuity*, Oxford, Blackwell, 1967.

SENTENCES AND NAMES IN FREGE

By H. W. NOONAN

IN his paper ' "*Bedeutung*" in Frege: A reply',[1] which is an answer to an earlier paper by Professor Ernst Tugendhat, V. H. Dudman writes as follows:

> Although the doctrine that sentences are names of objects was not announced by Frege until 'Function and Concept' (1891), it was implicit in his work from the beginning. For it is entailed by the treatment of sentences as completed functional expressions, and that treatment dates from *Begriffsschrift* (1879).

A little later he writes:

> I am not of course offering to defend the functional theory of predication or its corollary that sentences are names. I have no objection to Tugendhat's discarding these doctrines. (On the contrary, I share what I take it is the majority view, that, although the conception of the singular sentence as a completed functional expression is historically important, as having provoked the notion of an *open sentence* and hence provided the genesis of quantification theory, there remains, now that the innovation has established itself, no further motive for insisting that predicates form *singular terms* out of singular terms. The status of the functional theory of predication is that of a fruitful metaphor; nothing more.)

The assertion that the functional theory of predication[2] entails the doctrine that sentences are proper names is quite mistaken; in fact, as I shall now argue, there is no entailment either way.

The line of reasoning which led Frege to believe that sentences were proper names is easily reconstructed. He just assumed that all entities divide exhaustively into objects and functions.[3] Now sentences have a kind of completeness which prevents them from being functions, hence, given this assumption, they must be objects. But then their referents must also be objects, and so (since whatever stands for an object is a proper name) they must be proper names. Wittgenstein responded to this line of thought by denying Frege's assumption that all entities divide exhaustively into objects and functions; his ontology included

[1] ANALYSIS 33.1 (October 1972), pp. 21–27.

[2] The best accounts of the functional theory of predication are given by Geach and Dummett. For Geach's account see his essay on Frege in *Three Philosophers*, Basil Blackwell, Oxford, 1967, or his paper 'Names and Identity' in *Mind and Language*, ed. S. Guttenplan, Clarendon Press, Oxford, 1975. Dummett's account is given in his book *Frege: Philosophy of Language*, Duckworth, 1973. Geach accepts the functional theory of predication unreservedly, Dummett accepts it for what he calls 'complex' predicates, but not for what he calls 'simple' predicates.

[3] See 'Function and Concept' in Geach and Black (edd.), *Translations from the Philosophical Writings of Gottlob Frege*, Basil Blackwell, 1970, p. 32.

facts, and this, according to him, is what sentences were.[1] The assumption which licenses the move from sentences' being objects to their standing for objects, the assumption, namely, that what an expression signifies must be of the same logical category (object, function or fact) as the expression itself, was, of course, made by both Frege and Wittgenstein.

Now the essential premisses of this argument, considered as an argument for the doctrine that sentences are proper names of objects, are just these four: (a) that all complete entities are objects, (b) that sentences are complete entities, (c) that the referent of an object must itself be an object and (d) that an object which refers to another object must be a proper name of that object. But it is quite compatible with these premisses that concepts (the referents of predicates) should not be functions. The reason for this is that it is quite compatible with them that *incomplete* entities, about which they say nothing, should be divisible into various categories, only some of them being functions. One could thus (if one were perverse enough) accept Frege's assimilation of sentences to proper names while rejecting his functional theory of predication.

Consequently, it is worth noting, while the thesis that a sentence containing an empty proper name (i.e. analysable as formed by attachment of a predicate to that proper name) lacks a referent is entailed by the functional theory of predication, it is *not* entailed by the doctrine that sentences are proper names (since this doctrine does not entail that theory). This conflicts with the following statement in Dummett's book, p. 412.

> If sentences are agreed to have truth-values as their referents, but this is regarded, not as a special case, but merely as an analogue, of names having objects as their referents, then . . . there is no cogent argument from first principles to the conclusion that sentences containing a name which lacks a bearer are devoid of truth-value. The tenability of this conclusion, therefore, becomes a test for the coherence of Frege's later assimilation of sentences to complex names, since this assimilation certainly does entail the conclusion.

The functional theory of predication may not be a 'first principle' but it is, as Dudman says, the most distinctive feature of Frege's semantics; consequently, contrary to what Dummett here (and elsewhere throughout this chapter) implies, Frege's adherence to the thesis that a sentence containing an empty proper name lacks a referent cannot be regarded as an unfortunate error which can easily be corrected without modification of any of his central doctrines.

On the other hand, one could espouse the functional theory of predication while rejecting the assimilation of sentences to proper

[1] *Tractatus Logico-Philosophicus*, 3.14: 'A propositional sign is a fact'.

names. This is what Wittgenstein does: *Tractatus* 3.318 says 'Like Frege and Russell I construe a proposition as a function of the expressions contained in it.' It is only together with the assumption that only objects can be the values of functions that the functional theory of predication entails the doctrine that sentences are proper names, and this is an assumption for which Frege never gives any reason. For elaboration see Dummett's book, p. 249. Putting *Tractatus* 3.14 and 3.318 together one can see that Wittgenstein is committed to rejecting this assumption. Rejecting it leaves one with a choice: one can either hold that *neither* sentences *nor* what sentences stand for are objects or one can hold that while sentences are objects what they stand for are not. This latter course appears to be favoured by Dummett, but it involves rejecting the Frege-Wittgenstein view that what an object stands for must itself be an object.

Frege's functional theory of predication makes a claim about what kind of *incomplete* entities there are, it implies nothing about the classification of *complete* entities: his doctrine that sentences are proper names makes a claim about what kind of *complete* entities there are, it implies nothing about *incomplete* entities. The two are, therefore, quite independent.

Trinity Hall, Cambridge

PROMISSORY OBLIGATIONS AND RAWLS'S CONTRACTARIANISM

By Michael H. Robins

IN making a promise one is often taken to be creating a special obligation where none existed before, and hence promising is typically regarded as a paradigm of a moral obligation. Yet it is precisely such an analysis which seems so impossible; for how can a bona fide moral obligation be the kind of thing which is created out of thin air—either by an act of will or by a convention? That it would seem difficult to create by an act of will was argued effectively by Hume, who asked how it was even intelligible, much less plausible, for the will to obligate itself (*A Treatise of Human Nature,* ed. L. A. Selby-Bigge, Oxford: Clarendon Press, 1888, pp. 516–525). By an act of volition, Hume thought, one can intend an action, undertake a project, etc., yet such

undertakings are scarcely more than resolves, not promises. For they do not seem to give rise to an obligation as promising typically does. But Hume's most telling objection is this: even if one could make sense of such a thing as undertaking or willing an obligation, that is, make sense of it in the absence of human conventions or any act of communication, how is the obligation thus undertaken morally binding or "objectively valid", rather than an obligation which one merely *thinks* one is under? After all, if believing something doesn't make it so, how can 'willing' *create* an obligation?—the implication being that obligation must stand to volition as truth stands to belief.

Prichard expanded on Hume's doubts (in 'The Obligation to Keep a Promise', in his *Moral Obligation,* Oxford, 1929, pp. 169–179) by asking how it was possible even for a convention to create the obligation in the required way—the solution embraced by Hume (*Treatise,* pp. 519–525). For Prichard thought that an obligation just wasn't the sort of thing that could be created at all—at least not directly. Nevertheless, the explanation of promising by means of a convention is the most plausible candidate—and is so regarded today—for promising, as a public, communicative act, is a move in a game one of the rules of which requires that it be kept. Indeed Searle, who typifies this view, says that to ask how a promise can create an obligation is like asking how a touchdown can create six points (*Speech Acts,* Cambridge, 1969, p. 35; cf. pp. 33–42). Yet Prichard claims that this is impossible if we understand precisely the problem about creating an obligation. His doubts, if only because they are so timely, are worth quoting.

> Promising requires the actual use of the word 'promise' or some equivalent, such as 'undertake', 'agree'. . . . This being so, we can at least say that when I promise X to do some action, I am causing X to hear a certain noise, which has a definite meaning both to X and to me, together with the term ordinarily used for the action, in such a way that X believes that the sounds proceeds from me. But then the question at once arises: 'How can my doing this give rise to an obligation to do the action?' And the answer seems to have to be: 'Only because I have *already promised* not to cause a noise of that kind in connexion with the phrase for some action, without going on to do the action.' And if this answer is right, what we call promising to do some action appears [not to be *creating* an obligation in the performance of a speech act, but] to be causing someone to hear in connexion with the term for the action a noise of a kind which I *have promised* never to use in this kind of way without going on to do the action; and, again, the reason why, when I say to X 'I promise to do so and so', I am bound to do the action is that I have previously made him a general promise. (P. 172, Italics mine.)

Thus, according to Prichard, the invocation of a convention can give rise to a moral requirement to perform an action, not by, strictly speaking, creating that requirement, but only insofar as the convention itself invokes the terms of a prior and more general agreement by which that

convention was established. But, as Prichard recognizes, this solves the problem at one stage only to transmit it to the next; for unless this general promise is itself explained, he is on the threshold of an infinite regress. Thus he ends his essay with what is now a famous epigram:

> But, of course, it would be more accurate to say that what I am suggesting is not a conclusion but a problem for consideration; viz. what is that something implied in the existence of agreements which looks very much like an agreement and yet, strictly speaking, cannot be an agreement? (Ibid., p. 179.)

In this note I shall show how Rawls, interestingly enough, follows Prichard's account in its basic respects (in *A Theory of Justice,* Harvard, 1971, pp. 342–350)—something he himself acknowledges doing in part (p. 347)—but believes that Prichard's difficulties can be overcome by the hypothetical contractarian doctrine in connection with Rawls's 'principle of fairness'. I shall argue that in producing his attempted solution Rawls not only fails to stop the infinite regress, but undermines his contractarian position as a whole, inasmuch as he seems to acknowledge Prichard's difficulties about promising. For I will show that Rawls's contractarian position presupposes the concept of promising in such a way that the tenability of the whole position rests upon a solution to Prichard's difficulties.

Rawls begins by noting with Prichard that the institution or convention of promising does not by itself create bona fide moral obligations (pp. 345–46; 348–49). Indeed promises for Rawls are no different from any other institution whose rules define the content of the obligations within it. Thus, for example, the political obligations of a public official in a constitutional government are typically defined by the duties of the office and its place in the political scheme. In a broader sense, promising is regarded as being on a par with such disparate things as the use of political power and the deprivation of liberty, in the sense that all of these attain moral legitimacy ultimately only in virtue of a contractarian justification.

In any case, what, according to Rawls, is needed to turn internal institutional obligations, like the obligation to keep a promise, into binding moral obligations? Rawls's answer comes in two stages. The first is that a promise can be morally binding only if, *qua* institutionalized fact, it exemplifies a certain principle of fairness. The 'principle of fairness' in turn 'holds that a person is required to do his part as defined by an institution when the institution is just, that is, satisfies the two principles of justice,[1] and when he has voluntarily accepted the benefits

[1] 'First: each person is to have an equal right to the most extensive basic liberty compatible with a similar liberty for others' (p. 60). 'Second: social and economic inequalities are to be arranged so that they are both (a) to the greatest benefit to the least advantaged and (b) attached to offices and positions open to all under conditions of fair equality of opportunity' (p. 83).

of an arrangement or taken advantage of the opportunities it offers to further his own interest' (p. 111). The essential idea, as Rawls expounds it, is that 'we are not to gain from the cooperative labours of others without doing our fair share' (p. 112).

But how does the promiser—supposing his promise to be sincere—necessarily *benefit* from the institution of promising in such a way as to advance his own ends? This seems somewhat odd, because a promise as such, as opposed to a contract, is typically for the benefit of the promisee, not the promiser. I think that Rawls's reply to this is that for him the standard reason for invoking the institution on a particular occasion is to stabilize small schemes of cooperation. Accordingly, the most plausible answer to be extracted from this is that when one makes a promise one is typically inviting the promisee to *rely* on the performance of a future act in a way that he could not in the absence of anything less than a promise. If these expectations are disappointed, say, out of self interest, or even out of a slight inconvenience on the part of the promiser, then the promisee is put at a decisive disadvantage in relation to the promiser. In such a fashion the promiser advances his own ends only by making and then breaking a promise.

But this does not, of course, answer the question we were asking, which was how the promiser *qua* promiser necessarily benefits from the *keeping* of his promise (i.e., conforming to the institution) on a particular occasion. The situation described above is particularly seductive in that it intimates, though not nearly as much as do Rawls's actual remarks (on pp. 346–347), a typical conflict of interest situation in which a promise is a compromise solution. But this seems to be definitive, not of promises, but of contracts, in which the notion of striking a bargain is the *sine qua non*. (See Robert Samek, 'Performative Utterances and the Concept of Contract', *Australasian Journal of Philosophy*, 43, 1965, 198, 204–06, for a convincing exposition of the point developed here.) But a promise is not essentially bilateral in this fashion, but unilateral. This seems to suggest that at the very least they are distinct concepts—a point countenanced elsewhere by Rawls himself, in which passage he identifies two types of voluntary obligations supposedly governed by the principle of fairness: those that arise out of express or tacit undertakings, and those that arise out of voluntarily accepting benefits (p. 113). Yet in supposing that promissory obligations depend for their legitimacy upon both parties accepting benefits (via the principle of fairness) Rawls is either not distinguishing between promises and contracts (or, more generally, between obligations arising from undertakings and those arising from accepting benefits) or, what seems more likely, is assimilating the former to the latter, treating the latter as more basic.

We shall see how this relatively small point portends much greater difficulties for him at the more theoretical stage of the original position. For the issue in no small part will be whether the concept of a contract in the original position will be the more fundamental concept to explain promissory obligations or whether—as I shall maintain—such a contract will have to presuppose as more primitive the concept of promising, which is the very thing it is supposed to explain.

Let us now return to the main thread of the argument. The main thrust of the principle of fairness is that it is required as an extra premiss to enable us to derive from the fact that someone made a promise the obligation for him to keep it. The rationale for this is that 'the contract doctrine holds that no moral requirements follow from the existence of institutions alone. . . . Thus, along with most other ethical theories, justice as fairness holds that natural duties and obligations arise only in virtue of ethical principles' (p. 238).

But what makes those 'ethical principles' (in the case of the principle of fairness) themselves binding? The answer, which delineates the second stage, comes in the next sentence: 'These ethical principles are those that *would* (italics mine) be chosen in the original position' (ibid). When applied to the institution of promising Rawls's position comes out to be somewhat like that of Prichard, viz., that promises in the ordinary sense (defined by convention) cannot by themselves create a moral, fiduciary obligation, but can do so, ultimately, only in virtue of a general agreement to keep agreements. One main difference from Prichard's view, however, is that this agreement need not be a prior, actual (or tacit) agreement, but only a hypothetical agreement to commit ourselves to the principle of fairness.

Now I wish to point out that if, as Rawls thinks, promising is an ordinary institution on a par with those noted above, then there is something to be gained in the shift from the actual agreement to the hypothetical agreement. For such a shift marks the alleged general advantages of hypothetical contractarianism over its classical counterpart. The standard problem with classical contractarianism—exemplified in Prichard's agreement to keep agreements—is how such an agreement is to be understood as an historical event or, short of that, how it is to be construed as a form of tacit consent, voluntary acceptance or the like. Indeed for Rawls the more serious problem of the classical view of tacit consent is the threat it poses to the stability of institutions. For on any such account, he writes, 'citizens might still wonder about one another whether they were bound [i.e., in what ways they have manifested their 'voluntary acceptance'], or so regarded themselves. The public conviction that all are tied to just arrangements would be less firm, and a greater reliance on the coercive powers of the sovereign might be necessary to achieve stability' (p. 337). To adopt Prichard's

position, then, is at the very least, to be saddled with all of the standard problems of classical contractarianism.[1]

But while such a manoeuvre may successfully avoid the standard types of difficulties concerning classical contractarianism, it cannot work in the case of the so-called institution of promising because of promising's special relation to any form of contractarianism. The relationship, I will argue, is that *any* contractarian position—classical or hypothetical—needs to *presuppose* the validity of promissory obligations instead of, as Rawls implies, the other way around. In this way the institution of promising is emphatically unlike the other institutions which can plausibly be justified in a contractarian fashion. To illustrate how promising is presupposed as a primitive concept, consider Rawls's exposition of the agreement in the original position:

> There is one further assumption to guarantee strict compliance to the terms of the agreement. The parties are presumed to be capable of a sense of justice. . . . This condition is to ensure the integrity of the agreement made in the original position. . . . It means that the parties can rely on each other to understand and to act in accordance with whatever principles are finally agreed to. . . . They are rational in that they will not enter into agreements they know they cannot keep, or can do so only with great difficulty. (P. 145.)

A more trenchant passage occurs in the defence of the two principles of justice:

> Thus the parties consider the strains of commitment. . . . Since the original agreement is final and made in perpetuity, there is no second chance. In view of the serious nature of the possible consequences the question of the burden of the commitment is especially acute. . . . Moreover, when we enter into an agreement we must be able to honour it even should the worst possibilities prove to be the case. (P. 176.)

What these passages suggest is that we shall *assume*, for the purpose of the original agreement, that promises are morally obligatory and are so regarded by the parties to the agreement! (That is why we had better be very careful about what we commit ourselves to.) The difficulty should now be obvious. If promises as institutions cannot by themselves secure a moral obligation (because of Prichard's and Hume's objections concerning creating an obligation, the will obligating itself, etc.), then these same difficulties are not going to go away at the level of the original agreement to keep agreements—itself a kind of promise. It is precisely because they reappear that we are in the grip of a regress. Conversely, if despite these difficulties promising is miraculously able to generate bona fide moral obligations at the level of the original position, then this can

[1] I believe these so-called standard difficulties can be resolved by understanding the original promise as implied (logically) by certain explicit promises (in referentially opaque contexts) rather than as some psychological act of tacit consent. This point is developed, though inchoately, in my article, 'The Primacy of Promising', forthcoming in *Mind*, 85 (July, 1976).

also be the case with the promising in the ordinary and standard sense. (This is not to say, of course, that any promise, made under any sort of condition, can be binding, but only one made in the standard case. To assume, against Prichard and others, that promising can straightforwardly obligate is not to deny that either the concept or the corresponding oligation is defeasible.)

It might be objected that promising can obligate for Rawls in the original position but not in the ordinary case, because the original position characteristically has something that ordinary promises lack, namely, strict equality and fairness manifested in the veil of ignorance. This would seem to remove the inconsistency by making these characteristic features ultimately responsible for the creation of genuine promissory obligations (as indeed they would presumably be responsible for all of Rawls's moral principles). And since these features can be found only in the original agreement, they constitute the necessary and ultimate step in the contractarian justification of promissory obligations in the ordinary sense. Nor would such a procedure be in any way circular, because what is presupposed in the original agreement, viz., the veil of ignorance, etc., is an added, *sui generis,* feature not present in ordinary promises. In fact on this view the *promissory act* in the original position becomes a superfluous parable, obscuring the fact that it is the strict equality that carries the burden of justification. It is precisely because ordinary promises as such lack this equality (as well as the principle of fairness, which is different) that Rawls says that considerations of fairness in general are not part of the *concept* of promising, that the institution is not just by definition. These considerations are rather "external" to whatever institutional obligations are sanctioned by convention.

But it seems clear that the veil of ignorance cannot, without the promissory act, carry so heavy a burden. More generally, the pure procedural justice characterized by the original position needs the *act of promising*—however hypothetical—as an ineliminable part. The most convincing reason for this is that Rawls cannot generate from the original position the principles he wants, e.g., the two principles of justice, unless such principles are regarded as the outcome of a choice made under what he himself calls 'the strain of commitment'—that is, a choice in which each party gives up the right to change his mind (see p. 195). Only with such a background does it make sense to say, for example, that the parties could not 'run [the] chance of having to acquiesce in the loss of freedom over the course of their lives for the sake of the greater good enjoyed by others' (p. 176). Perhaps a more fundamental reason for the necessity of a binding agreement is that, given Rawls's view that moral principles are not necessary truths, the idea of an agreement made under conditions of equality seems to be a

reasonable alternative to scepticism. As this point is argued by Lyons (in 'The Nature of the Contract Argument', *Cornell Law Review*, 59, Aug., 1974, 1073–75), if we concede with the sceptic that we cannot *discover* valid moral principles by any independent criterion, then we can still provide a rational justification by showing that such principles are literally *created* by a procedure we agree is fair. But, of course, this act of creating principles not independently justified leads us inescapably to the presupposition that promise is at work, and with it to the now familiar difficulties concerning creating an obligation that were alleged of promises in the ordinary sense. But it was precisely such difficulties which in the present context gave rise to the need for the original agreement in the first place.

We are thus led to conclude that the original position requires the concept of promising as primitive—the promise being precisely something like an agreement to keep agreements. Contrary to Rawls's diagnosis, what makes the agreement mysterious is not merely the difficulties associated with classical contractarianism, but the necessity for it to *presuppose* that we can mysteriously bind the will in order to *explain* how it is possible to do such a thing. It has to presuppose a solution to the very difficulty to be elucidated. Now it should be obvious that Rawls's hypothetical agreement is not one whit more cogent on this score than is an actual agreement. For the mysterious ability to create moral obligations is being presupposed no less.

One can avoid this difficulty by taking the contractarian position to presuppose as a primitive concept that promising creates an obligation. The thrust of doing so is to concede that it cannot be explained by other notions, like original contracts, conventions, etc.; it is rather these notions and the like which must be explained in terms of promising. I have developed such a position in 'The Primacy of Promising', *op. cit.*, in which I argue that, *pace* Rawls, the obligation to keep a promise is a necessary, though non-analytic, truth, its justification being conceived along the lines of a transcendental argument. In such a fashion promissory obligations are claimed to be a necessary condition for the intelligibility of a whole panoply of rule-governed and conventional behaviour. Such a transcendental turn accomplishes much of what contractarianism—as commonly understood—seeks to accomplish: it admits with the sceptic that even though moral principles lack an independent justification, they are not arbitrary when regarded as self-imposed through basic agreements—real or implied—by autonomous persons. But unlike contractarianism, it traces the justificatory burden of the 'choice under the strain of commitment', to a fundamental first principle, which, as indicated, must be regarded as a necessary truth. The force of such a necessary truth is that the contractarian method of justification, if it is successfully to create binding moral principles, must

give way to the transcendental method of justification, and with it effect a re-assessment of the nature and role of necessary truths in ethics.

In the absence of such a manoeuvre, however, there are two conclusions which can be drawn about Rawls's hypothetical contractarianism. First, with regard to his treatment of promising, it appears to be inherently inconsistent. For his general agreement to keep agreements must presuppose the intelligibility and possibility of promising's creating a moral obligation, only to deny such a possibility at the level of ordinary agreements. He clearly cannot have it both ways. Secondly, and more importantly, the contractarian doctrine as a whole must presuppose as primitive (and as a necessary truth?) that promising can obligate; yet it cannot presuppose this, first, because of the difficulties Rawls acknowledges, which were raised by Hume and Prichard, concerning promising's creating an obligation,[1] and, more generally, because of its methodological assumptions regarding the unavailability of necessary truths to ethics (to which problem the contractarian doctrine is offered as a solution). The difficulty, then, is that the need to presuppose something inconsistent with its methodological and other assumptions seems to undermine as a whole the contractarian justification—as commonly conceived—in ethical theory.[2]

Bowling Green State University

[1] Note that these difficulties emerge only if the contractarian mode of justification does not give way to the transcendental mode of justification, outlined in the preceding paragraph.

[2] The germ of the idea of this paper grew out of a stimulating colloquium discussion as part of an N.E.H. Summer Seminar on Utilitarianism, by Richard Brandt (University of Michigan, 1975). I am very grateful to the members of that colloquium for their contribution: in particular, to Alan Fuchs (Rawls's most tenacious defender). I am also exceedingly grateful to two of my colleagues, Donald Scherer and Joan Leguard, for their incisive criticisms of an earlier draft of this paper.

PHILIPPA FOOT ON HYPOTHETICAL IMPERATIVES

By ROBERT L. HOLMES

MRS. FOOT'S rejoinder (ANALYSIS 35.2) to my critique (ANALYSIS 34.3) of her discussion of hypothetical imperatives bears out my original observation that her criticisms of the Kantian categorical imperative were misdirected. For she cites, as presumably representative of the position she is attacking, a quote by W. K. Frankena from G. C. Field to the effect that '. . . it is "one of the most deeply recognized characteristics of the moral fact" that it is in itself and necessarily "a reason for acting" '[1]; and then adds: 'This is what I am denying when I deny that moral judgments are categorical imperatives' (p. 53). But on *this* point Mrs. Foot and Kant are allies. For Kant, no less than Mrs. Foot, would deny that the established rightness or obligatoriness of an action in and of itself necessarily provides a 'reason for acting' in the sense of motivating the agent to perform the act (the sense involved in the Field quote). Moreover, the issue between those who assert that this is the case and those who deny it is not the issue of whether moral judgments are categorical imperatives, but is rather the issue between what Frankena calls 'externalists' and 'internalists' on the question of the relationship between obligation and motivation: internalists claiming (roughly) that obligation entails motivation, externalists (roughly) denying this. To deny the internalist position on this issue does nothing to show that moral judgments are hypothetical imperatives in any sense bearing upon the Kantian distinction between hypothetical and categorical imperatives.

Only slightly behind the categorical imperative in Mrs. Foot's low esteem is the concept of a justifying reason, though here, too, the precise nature of her objection to the concept is unclear. It is not as though she wants to hold on to motivating reasons in contrast to justifying reasons, for she seems to want to deny that very distinction, implying that it involves confusing (a) giving reason for a judgment with giving reason for an action, on the one hand, and (b) justifying a judgment with justifying an action, on the other. Whatever the bearing of these distinctions upon that between motivating and justifying reasons —and it is not at all clear why to make the latter distinction involves confusing the former—I did not, as she alleges, blur the distinction in (a), since it was precisely that distinction which I marked and found wanting in her own discussion of 'reasons for acting'. As for (b), I do maintain that to deny that to justify a judgment is to justify the action is incoherent, unless one assumes a shift in the frame of reference from

[1] W. K. Frankena, 'Obligation and Motivation', in A. I. Melden (ed.), *Essays in Moral Philosophy*: University of Washington Press, Seattle, 1958, p. 44.

which the justification is given. If one completely justifies the *judgment* 'I ought to do x' from a moral standpoint, there is nothing more that needs to be done to justify the *doing* of x from a moral standpoint—any more than, if one has shown the truth of the statement 'x is the case' there is more to be done to show that x *is* the case. So, if the claim that in morality one might be right if he denied that he had a reason to do what is moral means that once having justified the judgment 'I ought to do x' one can still rightly deny that he has a morally sufficient justification for doing x, the claim is clearly false. To justify the judgment, in this sense, is to justify the action. Whether to have done the latter is, in addition, to have given reasons for doing the action—another transition Mrs. Foot questions—depends, as I indicated, upon what one means by 'reason'. If one means considerations sufficient to move the person to perform the action, then to have justified an action does not (the internalist to the contrary notwithstanding) suffice to have given reasons for doing it. If one means considerations sufficient to justify the performance of the action, then obviously it does. Whether or not one objects to the labels 'motivating reasons' and 'justifying reasons' to mark this distinction, the distinction stands, and is essential to the understanding of the issue Mrs. Foot seeks to clarify. And we are left, in any event, without any philosophically interesting sense in which morality has been shown to be a system of hypothetical imperatives.

University of Rochester

ON *A PRIORI* CONTINGENCY

By DOUGLAS ODEGARD

IN 'On *A Priori* Contingent Truths' (ANALYSIS 36.2) W. R. Carter objects to Saul Kripke's view that if 'metre' is defined as the length of rod *S* (at t_0), then

(1) *S* is one metre long

is *a priori* though contingent. Carter argues that (1) is not *a priori* because it implies that *S* exists. But this tells us nothing about variants on (1) which avoid such an implication, like

(1*a*) Provided *S* exists, *S* is one metre long.

Are we to assume that if the proviso in (1*a*) is coherent, then (1*a*) *is a priori* and contingent?

If we let 'metre' refer to a (length equal to a) given length and identify the length as *S*'s length, then in order to know that something is a metre long we must know that it is (the same length as) the length which *S* is. Thus, to know that *S* is one metre long, we must know that *S* is the length which *S* is. If the given length is identified *accidentally* as *S*'s length, then *S* might have been a different length and hence not a metre long. In that case (1*a*) is contingent. But then our knowledge that *S* is the given length is not *a priori*. Granted, in defining 'metre' as the length which *S* happens to be, we know that *S* is that length. But this is not a consequence of the definition. Rather, it is a precondition. Thus, the fact that such a definition automatically yields knowledge that *S* is one metre long does not mean that the knowledge is *a priori*. The result is automatic because the knowledge that *S* is one metre long is non-linguistically identical with the knowledge that *S* is the given length, empirical knowledge which we must already have. Consequently, the knowledge that *S* is one metre long is not *a priori*. Kripke fails to see this when he argues: 'For if he used stick *S* to fix the reference of the term "one meter", then as a result of this kind of "definition" . . ., he knows automatically, without further investigation, that *S* is one meter long.' (Davidson and Harman (edd.), *Semantics of Natural Language*, Reidel: 1972, p. 275.) The knowledge is automatic because the individual must already know that *S* is the given length, not because the knowledge is *a priori*.

Alternatively, if we identify the given length *essentially* as *S*'s length, then our knowledge that *S* is that length is *a priori*. But this is because it is the knowledge that *S* is the length it is, whatever that length might be. In that case (1*a*) is equivalent to

(1*b*) Provided *S* exists, *S* is whatever length it is

and (1*b*) is necessary. We cannot say that *S* might have been a different length, since any purportedly different length of *S* now counts as "the length which *S* is" and hence is not different. Granted, *S* might have been a different length in the sense that, if *S*'s length is *n* non-metric units, then *S* might not have been *n* units long. But this does not mean that *S* might have been such that it is not whatever length it is. There are possible worlds in which *S* is not *n* units long. But there is no possible world in which *S* is not the length it is *if* 'the length it is' is used essentially.

Nor can we create an *a priori* contingency by replacing (1*b*) by

(1*c*) Provided *S* exists, *S* is, as a matter of contingent fact, the length it is.

The only new factor in (1*c*) is the implication that *S* might not have been the length it is, But an assertion of possibility is itself necessary. In modal logic: $Mp \rightarrow NMp$. In possible worlds language: if there is a possible world in which *S* is not *n* units long, then in every possible world there is a possible world in which *S* is not *n* units long. Hugh S. Chandler, in 'Plantinga and the Contingently Possible' (ANALYSIS 36.2), rejects such a principle (though not with a view to establishing that something like (1*c*) is contingent *a priori*). He feels that, although a bicycle can survive the gradual replacement of all its parts, a given bicycle could not have existed at a given time with a completely different set of parts. If we allow this, then we must allow (i) that an individual x which is *ab* in world w_0 could have been *ac* but not *cd*; and (ii) that if in world w_1 x had been *ac*, then x could have been *cd*. In that case, in w_1 there is a possible world in which x is *cd*, but in w_0 no such world is possible. But why should we grant the starting-point? Trans-world identity and identity through time seem to be on a par. If the bicycle *can* be completely re-composed through time, then it *could* have existed at any given time with every part different. If the latter is impossible, then so is the former. Now, the assumption that re-composition must result from a gradual process might seem a stumbling block. But in the first place, it is not clear that non-gradual re-composition is always ruled out *logically*; it may simply be incompatible with our *telling* that the individual persists. And even if it is excluded logically, this only means that if the bicycle had entirely different parts at a given time, it could not have just previously had the same parts. It does not mean that the bicycle could not have had entirely different parts. Chandler's remarks therefore contain nothing to support the contingency of (1*c*).

In specific contexts, of course, the assertion '*S* is the length it is' can be contingent; *e.g.* when it is elliptical for

(1*d*) Provided *S* exists, *S* is the length it is, *viz.* *n* units.

But such contexts fail to yield anything *a priori*.

For these reasons, I think that even if variants on (1) avoid existential commitments, they still cannot be both *a priori* and contingent.

University of Guelph

A FORMALIZATION OF GEACH'S ANTINOMY

By JOHN DAVID STONE

THE line of argument formalized below was invented by P. T. Geach and presented in a note entitled 'On *insolubilia*' (ANALYSIS 15.2, reprinted in P. T. Geach, *Logic Matters*, Oxford, Basil Blackwell, 1972). A comparison of Geach's version with the antinomy presented here may make clearer the relation between natural languages (English, in the case of Geach's *insolubile*) and formal, artificial ones.

The following formal language, arbitrarily named Stub, has deliberately been kept as simple as possible. It contains only what is necessary for the construction of the antinomy. (Naturally, the same antinomy occurs in any extension of Stub produced by adding to it a more powerful class of axioms.) Because the antinomy occurs in Stub, the language is worthless as an instrument of discourse; but it is useful as a means of exhibiting the antinomy.

Syntax

Alphabet: $x, y, Q, -, +, T, =, \rightarrow, \{, \}, <, >, [,]$.

Syntactic rules:

1. x and y are terms.
2. For any string μ, $Q\{\mu\}$ is a term.
3. For any terms τ, υ, and ϕ, $<\tau-\upsilon+\phi>$ is a term.
4. For any term τ, $T\tau$ is a sentence.
5. For any terms τ and υ, $[\tau = \upsilon]$ is a sentence.
6. For any sentences ρ and σ, $[\rho\rightarrow\sigma]$ is a sentence.
7. No other string is well-formed.

Logic

Axiom rules:

1. For any strings λ and μ, $[<Q\{\lambda\}-y+Q\{\mu\}> = Q\{\varkappa\}]$ is an axiom, where $\varkappa$ is the result of replacing every x in λ by μ.

2. For any term τ and any sentence σ, $[[\tau = \mathcal{Q}\{\sigma\}] \rightarrow [T\tau \rightarrow \sigma]]$ is an axiom.
3. For any term τ and any sentence σ, $[[\tau = \mathcal{Q}\{\sigma\}] \rightarrow [\sigma \rightarrow T\tau]]$ is an axiom.
4. There are no other axioms.

Deduction rules:

1. For any sentences ρ and σ, σ is immediately deducible from the class $\{\rho, [\rho \rightarrow \sigma]\}$.
2. For any sentences ρ and σ, $[\rho \rightarrow \sigma]$ is immediately deducible from the class $\{[\rho \rightarrow [\rho \rightarrow \sigma]]\}$.
3. The preceding clauses exhaust the relation of immediate deducibility.

Semantics

A function f is an interpretation of Stub exactly if it meets all of the following conditions:

C1. The domain of f is the class of well-formed strings of Stub.
C2. $f(y)$ is x.
C3. For any string μ, $f(\mathcal{Q})\{\mu\})$ is μ.
C4. For any terms τ, υ and ϕ, if $f(\tau)$, $f(\upsilon)$, and $f(\phi)$ are strings, $f(<\tau - \upsilon + \phi>)$ is the result of replacing every occurrence of $f(\upsilon)$ in $f(\tau)$ by $f(\phi)$.
C5. For any term τ, $f(T\tau)$ is 1 if $f(\tau)$ is a sentence and $f(f(\tau))$ is 1; otherwise $f(T\tau)$ is 0.
C6. For any terms τ and υ, $f([\tau = \upsilon])$ is 1 if $f(\tau)$ is $f(\upsilon)$; otherwise $f([\tau = \upsilon])$ is 0.
C7. For any sentences ρ and σ, $f([\rho \rightarrow \sigma])$ is 0 $f(\rho)$ is 1 and $f(0)$ is 0; otherwise $f([\rho \rightarrow \sigma])$ is 1.

The designated truth-value is 1.

A few explanatory remarks concerning conditions C2 through C7 are in order. C2 requires that y denote x. In the antinomy it will be necessary to have a name for x that does not itself contain x, and this is the simplest way to arrange that. From C3, the role of the operator $\mathcal{Q}$ is apparent; it is an equivalent of single quotation marks, and of the phrase 'quotation of' in Geach's original formulation. C4 indicates how the operation of substitution within strings is represented in the language. As C5 shows, the predicate T is intended as an equivalent of 'is a true sentence'. C6 is the semantic definition of the identity sign, and also provides a way of saying within Stub that one string μ denotes another string λ: For any interpretation f, $[\mu = \mathcal{Q}\{\lambda\}]$ is true exactly if $f(\mu)$ is λ, that is, exactly if μ denotes λ. Finally, C7 defines the connective $\rightarrow$

as standing for ordinary material implication. This is usually symbolized by a horseshoe; here the arrow was chosen in order to emphasize that the connective need not be defined in this way for the antinomy to occur. All that is necessary is that all instances of the axiom rules be valid and that the deduction rules never permit the deduction of an invalid sentence from a class of valid ones. Thus the arrow might equally well stand for entailment in a modal system, or any of several other things.

Given this semantics, the axioms seem fairly plausible. In axiom rule 1, the sentence schema is the Stub equivalent of the condition which follows it: $\varkappa$ is the result of replacing every $\varkappa$ in λ by μ. Axiom rules 2 and 3 make explicit the relevant assumptions in what Geach calls 'the naive view of truth': together they are simply a version of the Tarski biconditional schema,

$$\ulcorner \ldots \text{ is true exactly if } \text{———}. \urcorner$$

where the dash is to be replaced by a sentence and the dots by a term which denotes that sentence. The antecedent part of the sentence schemata in axiom rules 2 and 3 represents the stipulation about denotation; the consequents are the two halves of the biconditional.

In terms of this semantics, one can justify the logical rules and hence prove that Stub is sound, as follows:

(Axiom rule 1.) Consider any strings λ and μ. For any interpretation f, $f(\mathcal{Q}\{\varkappa\})$ is $\varkappa$, by C3. Hence, by C6, $f([<\mathcal{Q}\{\lambda\}-y+\mathcal{Q}\{\mu\}> = \mathcal{Q}\{\varkappa\}])$ is 1 exactly if $f(<\mathcal{Q}\{\lambda\}-y+\mathcal{Q}\{\mu\}>)$ is $\varkappa$. Now, since $f(\mathcal{Q}\{\lambda\})$, $f(y)$, and $f(\mathcal{Q}\{\mu\})$ are all strings—namely λ, $\varkappa$, and μ respectively—$f(<\mathcal{Q}\{\lambda\}-y+\mathcal{Q}\{\mu\}>)$ is, according to C4, the result of replacing every occurrence of $\varkappa$ in λ by μ. This, however, is exactly how $\varkappa$ is described in axiom rule 1. Thus, $f([<\mathcal{Q}\{\lambda\}-y+\mathcal{Q}\{\mu\}> = \mathcal{Q}\{\varkappa\}])$ is indeed 1, and so every axiom under axiom rule 1 is valid.

(Axiom rules 2 and 3.) Consider any term τ and any sentence σ. For any interpretation f, either $f(\tau)$ is σ or else it is not. If it is, then, since $f(\tau)$ is a sentence, it follows by C5 that $f(T\tau)$ is 1 if $f(\sigma)$ is 1 and 0 if $f(\sigma)$ is 0. In either case, both $f([T\tau\rightarrow\sigma])$ and $f([\sigma\rightarrow T\tau])$ are 1, by C7, and hence $f([[\tau = \mathcal{Q}\{\sigma\}]\rightarrow[T\tau\rightarrow\sigma]])$ and $f([[\tau = \mathcal{Q}\{\sigma\}]\rightarrow[\sigma\rightarrow T\tau]])$ are 1, again by C7. On the other hand, if $f(\tau)$ is not σ, then, since by C3 σ is $f(\mathcal{Q}\{\sigma\})$, $f(\tau)$ is not $f(\mathcal{Q}\{\sigma\})$. Therefore, by C6, $f([\tau = \mathcal{Q}\{\sigma\}])$ is 0. In this case too, then, $f([[\tau = \mathcal{Q}\{\sigma\}]\rightarrow[T\tau\rightarrow\sigma]])$ and $f([[\tau = \mathcal{Q}\{\sigma\}]\rightarrow[\sigma\rightarrow T\tau]])$ are 1, by C7. Thus any instance of axiom rule 2 or 3 is valid.

(Deduction rule 1.) For any sentences ρ and σ, if ρ and $[\rho\rightarrow\sigma]$ are valid, then in every interpretation f, $f(\rho)$ and $f([\rho\rightarrow\sigma])$ are both 1. It follows by C7 that $f(\sigma)$ is not 0; since the only other truth-value is 1, $f(\sigma)$ is 1. This is so for every interpretation, and hence σ is valid. There-

fore, no invalid sentence can be deduced from valid ones by means of deduction rule 1.

(Deduction rule 2.) For any sentences ρ and σ, if $[\rho\rightarrow[\rho\rightarrow\sigma]]$ is valid, then in every interpretation f, $f([\rho\rightarrow[\rho\rightarrow\sigma]])$ is 1. If now $f([\rho\rightarrow\sigma])$ were 0, then by C7 $f(\rho)$ would be 1, in which case $f([\rho\rightarrow[\rho\rightarrow\sigma]])$ would be 0, again by C7. Hence, $f([\rho\rightarrow\sigma])$ is not 0; it can only be 1. Thus $[\rho\rightarrow\sigma]$ is valid. So deduction rule 2 never leads from a valid sentence to an invalid one.

Every theorem of Stub is deduced from the axioms by means of the deduction rules. Since the axioms are valid and the deduction rules preserve validity, every theorem of Stub is valid. Thus Stub is sound.

Now, there are sentences of Stub which are not true in any interpretation of Stub. One of them is Ty. In any interpretation f, $f(y)$ is x, by C2; since x is not a sentence, $f(Ty)$ is 0, by C5. Surely, then, Ty should not be a theorem of Stub. Yet it can be deduced from the axioms by the deduction rules; it is this deduction which is the formal analogue of Geach's *insolubile*.

The deduction contains some ponderous sentences, which it will be useful to abbreviate. In naming the sentences which make up the antinomy, I shall write a delta in each of the places which would, but for human frailty, be filled by

$$<\mathcal{Q}\{[T<x-y+\mathcal{Q}\{x\}>\rightarrow Ty]\}-y+\mathcal{Q}\{\mathcal{Q}\{[T<x-y+\mathcal{Q}\{x\}>\rightarrow Ty]\}\}>$$

The important point about Δ is that, in any interpretation f, $f(\Delta)$ is $[T\Delta\rightarrow Ty]$. Confirming this is not as daunting a task as it might appear. Referring to C4, and noting that $f(\mathcal{Q}\{[T<x-y+\mathcal{Q}\{x\}>\rightarrow Ty]\})$ is $[T<x-y+\mathcal{Q}\{x\}>\rightarrow Ty]$, $f(y)$ is x, and $f(\mathcal{Q}\{\mathcal{Q}\{[T<x-y+\mathcal{Q}\{x\}>\rightarrow Ty]\}\})$ is $\mathcal{Q}\{[T<x-y+\mathcal{Q}\{x\}>\rightarrow Ty]\}$, one mechanically substitutes the third of these strings for the second throughout the first. The result is $[T\Delta\rightarrow Ty]$. The same operation establishes that $[\Delta=\mathcal{Q}\{[T\Delta\rightarrow Ty]\}]$ is an instance of axiom rule 1. The antinomy, then, is as follows:

1. $[\Delta=\mathcal{Q}\{[T\Delta\rightarrow Ty]\}]$	(axiom rule 1)
2. $[[\Delta=\mathcal{Q}\{[T\Delta\rightarrow Ty]\}]\rightarrow[T\Delta\rightarrow[T\Delta\rightarrow Ty]]]$	(axiom rule 2)
3. $[T\Delta\rightarrow[T\Delta\rightarrow Ty]]$	(1, 2, deduction rule 1)
4. $[T\Delta\rightarrow Ty]$	(3, deduction rule 2)
5. $[[\Delta=\mathcal{Q}\{[T\Delta\rightarrow Ty]\}]\rightarrow[[T\Delta\rightarrow Ty]\rightarrow T\Delta]]$	(axiom rule 3)
6. $[[T\Delta\rightarrow Ty]\rightarrow T\Delta]$	(1, 5, deduction rule 1)
7. $T\Delta$	(4, 6, deduction rule 1)
8. Ty	(7, 4, deduction rule 1)

Thus there must be some error in the preceding argumentation: Either Stub is not sound, or Ty is valid after all, for it certainly is a theorem. The resolution of this puzzle lies in the fact that there are no interpretations of Stub, because conditions C1 through C7 are incon-

sistent; no function meets all of them. Thus *Ty is* valid, vacuously, even though there are no interpretations in which it is true, since there are no interpretations in which it is false either.

The condition in which the inconsistency is centred is C_5, which does not in fact determine any value for $f(T\Delta)$. It stipulates that $f(T\Delta)$ is 1 exactly if $f(\Delta)$ is a sentence of which the truth-value is 1. But $f(\Delta)$ is $[T\Delta \rightarrow Ty]$, and the truth-value of this sentence depends in turn on that of $T\Delta$. Since if there were an interpretation f, $f(Ty)$ would be 0 by C_2 and C_5, it is clear from C_7 that $f([T\Delta \rightarrow Ty])$ would have to be 1 if $f(T\Delta)$ were 0 and 0 if $f(T\Delta)$ were 1. Each alternative is inconsistent with the original application of C_5.

The University of Texas at Austin

BERNAYS'S NON-CIRCULAR PROOF OF THE NON-INDEPENDENCE OF THE FOURTH AXIOM OF *PRINCIPIA MATHEMATICA*

By JAMES M. BROWN

AS it stands, Bernays's nineteen-line proof of PM's fourth axiom from the other axioms has gaps in it. (See pp. 312f. of P. Bernays: 'Axiomatische Untersuchungen des Aussagen-Kalküls der *Principia Mathematica*', *Mathematische Zeitschrift*, vol. 25, 1926, pp. 305–320.) J. Danquah (ANALYSIS 36.2, pp. 110f.) draws attention to this, and points out that we can fill the gaps by making use of *Syll* 2, viz.

$$p \supset q : \supset : q \supset r . \supset . p \supset r$$

Since the PM proof of *Syll* 2 depends on *Asso* (i.e. the fourth axiom), this makes the proof of *Asso* circular unless some other proof of *Syll* 2 is forthcoming. (In the last line of Danquah's note, '*Perm*' should presumably read '*Comm*'.)

However, we can alternatively fill the gaps by making use of *Syll* 1, viz.

$$q \supset r : \supset : p \supset q . \supset . p \supset r$$

and the unabbreviated proof is then not circular. In view of this, even if Bernays had supposed that *Syll* 2 was to be used in supplying the steps missing from an abbreviated proof, it would have been appropriate to attribute an oversight to the prover rather than to attribute circularity to the proof.

But Bernays did not suppose this. Before giving the proof, he made clear his use of *Syll* 1 in the abbreviated presentation of proofs. And he pointed out that *Syll* 1 is obtainable immediately from the fifth axiom: 'From the primitive formula *Sum* we obtain, as already mentioned, by substitution the formula

Syll: $q \supset r. \supset :p \supset q. \supset . p \supset r.$' (Op. cit., p. 311.)

What was 'already mentioned' (on p.309) is that *Simp* (i.e. $q \supset . p \supset q$) is a special case of *Add* (the second axiom) and *Syll* is the corresponding special case of *Sum*. These special cases are obtained by substituting '$\sim p$' for 'p' (and abbreviating by means of the definition of '$\supset$').

What might possibly be misleading is that Bernays set out the schema of the derived rule of inference for abbreviating proofs as:

$$\frac{\begin{array}{c} A \supset B \\ B \supset C \end{array}}{A \supset C}$$

But the order of the premises is quite immaterial.

I give one example to indicate how, in the unabbreviated proof, the gaps would be filled. From $p \supset r \vee p$ and $r \vee p \supset p \vee r$, Bernays obtains $p \supset p \vee r$. Danquah suggests, in effect, that the key missing formula is

$$p \supset r \vee p : \supset : r \vee p \supset p \vee r. \supset . p \supset p \vee r$$

Actually, it is

$$r \vee p \supset p \vee r : \supset : p \supset r \vee p. \supset . p \supset p \vee r$$

New University of Ulster